CU6-H05950-23-EP-2686

Calcutta University Library
ALIPORE CAMPUS

THE JOURNAL OF PHILOSOPHY

㉓

VOLUME CIV, NUMBER 1
JANUARY 2007

㉓

page		
5	*Understanding the Embodiment of Perception*	Kenneth Aizawa
26	*Macroscopic Mixtures*	Paul Needham
53	MEMO TO AUTHORS	
54	NEW BOOKS ANTHOLOGIES	

Published by **The Journal of Philosophy, Inc.**

THE JOURNAL OF PHILOSOPHY

FOUNDED BY FREDERICK J E WOODBRIDGE AND WENDELL T BUSH

Purpose To publish philosophical articles of current interest and encourage the interchange of ideas, especially the exploration of the borderline between philosophy and other disciplines

Editors Bernard Berofsky, Akeel Bilgrami, John Collins, Arthur C Danto, Kent Greenawalt, Patricia Kitcher, Philip Kitcher, Isaac Levi, Wolfgang Mann, Mary Mothersill, Frederick Neuhouser, Christopher Peacocke, Carol Rovane, Achille C Varzi *Consulting Editors* David Albert, James T Higginbotham, Robert May, Charles D Parsons, Wilfried Sieg *Managing Editor* John Smylie

THE JOURNAL OF PHILOSOPHY is owned and published by the Journal of Philosophy, Inc *President*, Arthur C Danto, *Vice President*, Akeel Bilgrami, *Secretary*, Daniel Shapiro, *Treasurer*, Barbara Gimbel, *Other Trustees* Lee Bollinger, Leigh S Cauman, Kent Greenawalt, Michael J Mooney, Lynn Nesbit

All communications to the Editors and Trustees and all manuscripts may be sent to John Smylie, Managing Editor, Mail Code 4972, 1150 Amsterdam Avenue, Columbia University, New York, New York 10027 FAX (212) 932-3721

You may also visit our website at www journalofphilosophy org

THE JOURNAL OF PHILOSOPHY

2007

SUBSCRIPTIONS (12 issues)

Individuals	$45 00
Libraries and Institutions	$100 00
Libraries and Institutions (print and online)	$180 00
Students, retired/unemployed philosophers	$20 00
Postage outside the U S	$15 00

Payments only in U S currency on a U S bank All back volumes and separate issues available back to 1904 Please inquire for advertising rates, ad space is limited, so ad reservations are required

Published monthly as of January 1977, typeset by Dartmouth Journal Services, Waterbury, VT, and printed by The Sheridan Press, Hanover, PA.

All communication about subscriptions and advertisements may be sent to Pamela Ward, Business Manager, Mail Code 4972, 1150 Amsterdam Avenue, Columbia University, New York, NY 10027 (212) 866-1742

All materials published in *The Journal of Philosophy* are copyrighted by the *Journal* and are protected by U S copyright law All rights reserved For more information about our copyright policies, please consult our January issue or contact us

The Journal of Philosophy (ISSN 0022-362X) is published monthly by The Journal of Philosophy, Inc , MC4972, Columbia University, 1150 Amsterdam Avenue, New York, NY 10027 Periodicals postage paid at New York, NY, and other mailing offices

POSTMASTER. Please send address changes to *The Journal of Philosophy* at MC4972, Columbia University, 1150 Amsterdam Avenue, New York, NY 10027

© Copyright 2007 by the Journal of Philosophy, Inc ISSN 0022–362X

Calcutta University Library
ALIPORE
CAMPUS

THE JOURNAL OF PHILOSOPHY

THE JOURNAL OF PHILOSOPHY was founded in 1904 as THE JOURNAL OF PHILOSOPHY, PSYCHOLOGY AND SCIENTIFIC METHODS by Frederick J E Woodbridge and J McKeen Cattell In 1906, Wendell T Bush became associated with the JOURNAL as co-editor The title was changed in 1921, and the JOURNAL was incorporated in the State of New York under its present name in 1923 Since its founding, it has been published from the Department of Philosophy at Columbia University It is typeset by Dartmouth Journal Services, Waterbury, VT, and printed by The Sheridan Press, Hanover, PA

KP 2686

THE JOURNAL OF
PHILOSOPHY

VOLUME CIV, NUMBER 1
JANUARY–DECEMBER 2007

Edited by
Bernard Berofsky, Akeel Bilgrami
John Collins, Arthur C. Danto
Kent Greenawalt, Patricia Kitcher
Philip Kitcher, Isaac Levi, Wolfgang Mann
Mary Mothersill, Christopher Peacocke
Carol Rovane, Achille C. Varzi

Consulting Editors
David Albert, James T. Higginbotham, Robert May,
Charles D. Parsons, Wilfried Sieg

Managing Editor
John Smylie

Published by The Journal of Philosophy, Inc.

STATEMENT OF COPYRIGHT POLICY

We would like to draw your attention to the protection copyright affords against multiple photocopying by libraries, instructors, and copy centers The 1976 copyright law provides that photocopies in excess of "fair use" copying cannot be made without authorization by the copyright owner—"fair use" meaning single copying for personal use Accordingly, the JOURNAL authorizes personal or educational multiple copying only if a request is made in writing, and the fee of $ 05 per page/per copy is paid directly to us *Each copy must include a notice of copyright* We shall turn over to the authors any substantial payments (over $30 00, an administrative fee) obtained from all photocopying Inquiries about photocopying may be sent to the Managing Editor Please consult our website, www journalofphilosophy org, for more information about our copyright policies

© Copyright 2007 by the Journal of Philosophy, Inc ISSN 0022-362X

POSTMASTER Second class postage paid at New York, NY, and other mailing offices Please send address changes to THE JOURNAL OF PHILOSOPHY, 709 Philosophy Hall, Mail Code 4972, 1150 Amsterdam Avenue, Columbia University, New York, NY 10027 Fax (212) 932-3721

KP 2686

THE JOURNAL OF PHILOSOPHY
VOLUME CIV, NO. 1, JANUARY 2007

UNDERSTANDING THE EMBODIMENT OF PERCEPTION

Obviously perception is embodied. After all, if creatures were entirely disembodied, how could physical processes in the environment, such as the propagation of light or sound, be transduced into a neurobiological currency capable of generating experience? Is there, however, any deeper, more subtle sense in which perception is embodied? Perhaps Alva Noë's theory of enactive perception provides one proposal.[1] Noë suggests a radical constitutive hypothesis according to which

> (COH) Perceptual experiences are constituted, in part, by the exercise of sensorimotor skills.

On Noë's view, bodily processes form a constitutive element in perceptual experiences. By contrast, it is more commonly supposed that bodily processes have at most a causal role to play in the genesis of perceptual experiences. Roughly stated,

> (CAH) Perceptual experiences are caused, in part, by the exercise of sensorimotor skills.

Clearly these two hypotheses offer distinct conceptions of the embodiment of perception. One view maintains that there is a constitutive dependency between bodily processes and perceptual experiences, where the other maintains that there is a causal dependency. This paper will make the case that Noë fails to offer any evidence that supports (COH) over (CAH) and that there is experimental evidence that favors (CAH) over (COH).[2]

---

[1] Noë, *Action in Perception* (Cambridge: MIT, 2004).

[2] An expanded version of this discussion will be found in Fred Adams and Ken Aizawa, *The Bounds of Cognition* (Malden, MA: Blackwell, forthcoming).

0022-362X/07/0401/5–25

I. SOME CONTEXT FOR THIS ISSUE

At first glance, a dispute over (CAH) versus (COH) may seem to be of limited interest or significance. Yet, it is easy to connect (COH) with the theory of extended cognition in ways that transcend Noë's interests or commitments.[3] On the plausible assumption that perceptual experience is the product of a cognitive process, (COH) apparently commits one to the view that at least some cognitive processes are constituted, in part, by the exercise of sensorimotor skills. On the further assumption that instances of the exercise of sensorimotor skills are constituted, in part, by processes in bodily components, such as peripheral nerves and muscles, we have it that at least some cognitive processes are constituted, in part, by bodily processes. This last, however, is part of the extended cognition hypothesis. As Tim van Gelder and Robert Port state the view, "Cognitive processes span the brain, the body, and the environment" (*op. cit.*, p. ix). Or, as Andy Clark and David Chalmers put it, "Cognitive processes ain't (all) in the head!" (*op. cit.*, p. 8). Or, as Robert Wilson states the view, "many cognitive capacities in symbol-using creatures, far from being purely internal, are either *enactive bodily* capacities, or *world-involving* capacities. These capacities are not realized by some internal arrangement of the brain or central nervous system, but by embodied states of the whole person, or by the wide system that includes (parts of) the brain as a proper part" (*op. cit.*, p. 188). The dispute between (CAH) and (COH) is, thus, a special case of the difference between the more general standard hypothesis of cognitive processing being causally connected to bodily and environmental processes and the more general radical (extended cognition) hypothesis that cognitive processing is constituted, in part, by bodily and environmental processes. If (COH) can be shown to be empirically confirmed over (CAH), this provides a beach head for the further advance of extended cognition. But, if (CAH) can be shown to be superior to (COH), the rising tide of extended cognition might be slowed.

---

[3] Some prominent contributions to the extended cognition literature are John Haugeland, "Mind Embodied and Embedded," in Haugeland, ed., *Having Thought* (Cambridge: Harvard, 1998), pp. 207–37; Andy Clark, "Reasons, Robots, and the Extended Mind," *Mind and Language*, XVI (2001): 121–45; Clark and David Chalmers, "The Extended Mind," *Analysis*, LVIII (1998): 7–19; Mark Rowlands, *The Body in Mind* (New York: Cambridge, 1999); Larry Shapiro, *The Mind Incarnate* (Cambridge: MIT, 2004); Tim van Gelder, "What Might Cognition Be, If Not Computation?" this JOURNAL, XCI, 7 (July 1995): 345–81; van Gelder and Robert Port, "Preface," in Port and van Gelder, eds., *Mind as Motion: Explorations in the Dynamics of Cognition* (Cambridge: MIT, 1995), pp. vii–x; and Robert Wilson, *The Boundaries of the Mind: The Individual in the Fragile Sciences* (New York: Cambridge, 2004).

Another reason to attend to the difference between (CAH) and (COH) is that it makes it clear that the plausibility of (CAH) cannot readily be used to support (COH) Why think that the causal interactions between the exercise of sensorimotor skills and cognitive processes is a basis for thinking that the exercise of sensorimotor skills constitutes, in part, cognitive processes? After all, the causal interactions between the expansion of a bimetallic strip in a thermostat and the cooling of an air conditioned room is no basis for thinking that the expansion, in part, constitutes the cooling, or that the cooling constitutes, in part, the expansion That we need to emphasize the distinction between hypotheses like (CAH) and (COH) is born out by a review of the literature discussing extended cognition, where shifts between something like them are made quite casually To take a few examples, Wilson describes playing a children's puzzle game, wherein one moves wooden rectangles around in a wooden frame,

> We solve the problem by continually looking back to the board and trying to figure out sequences of moves that will get us closer to our goal, all the time exploiting the structure of the environment through continual interaction with it We look, we think, we move But the thinking, the cognitive part of solving the problem, is not squirreled away inside us, wedged between the looking and the moving, but developed and made possible through these interactions with the board (*op cit*, p 194)

This looks to be a move from the cognitive processes causally interacting with the environment to the conclusion that cognitive processes are constituted, in part, by the environment Raymond Gibbs,[4] dealing with intentions, also seems to be up to the same thing

> The windsurfer continually affects and is affected by the set of the rig, so the behavioral intention to successfully windsurf emerges as a result of the interaction between the person and environment Focusing on the agent alone, or on how the agent responds to the environment, fails to capture the complex nuances of windsurfing behavior Just as it is important to understand the significance of paper and pencil when one does long division, where the cognition of doing long division is in part "offloaded" into the environment, the intentionality in windsurfing is best understood as a distributed cognitive behavior involving a person, a device, and the environment (*ibid*, pp 117–18)

---

[4] Gibbs, "Intentions as Emergent Products of Social Interactions," in Bertram F Malle, Louis J Moses, and Dare A. Baldwin, eds , *Intentions and Intentionality* (Cambridge MIT, 2001), pp 105–22

Noë, for his part, entertains, if not outright endorses, the same shift in the following passage,

> Some cognitive states—for example, states of thinking, calculating, navigating—may be partially external because, at least sometimes, these states depend on the use of symbols and artifacts that are outside the body. Maps, signs, and writing implements may sometimes be as inextricably bound up with the workings of cognition as neural structures or internally realized symbols (if there are any). According to active externalism, the environment can drive and so partially constitute cognitive processes (*op. cit.*, pp. 220–21).[5]

These examples are noteworthy, primarily for their brevity. Other examples of the insensitivity to the difference between causal connection and constitution may be found in Haugeland (*op. cit.*, p. 217), Clark and Chalmers (*op. cit.*, p. 8), Clark (*op. cit.*, p. 132), Shapiro (*op. cit.*, pp. 212–13), and Wilson (*op. cit.*, p. 195). So, (CAH) and (COH) merit attention because they are instances of more general hypotheses concerning the spatial extent of cognition and because they reveal more general problems with familiar arguments for extended cognition.

The foregoing analysis is supported by an emerging consensus among critics of the extended cognition perspective and of Noë's enactive theory of perception in particular. Fred Adams and Ken Aizawa[6] and Robert Rupert[7] discuss the problematic shift from the observation of causal dependencies to the conclusion of constitutive dependencies as it is found in the extended cognition literature. The issue also figures prominently in Ned Block's[8] and Jessie Prinz's[9] reviews of Noë's book. The first section of Block's review, in particular, is entirely consonant with the arguments developed here. It does not, however, directly address Noë's arguments for

---

[5] This kind of argument is discussed in greater detail in two works by Adams and Aizawa, *The Bounds of Cognition*, and "Why the Mind Is Still in the Head," in Philip Robbins and Murat Aydede, eds., *Cambridge Handbook of Situated Cognition* (New York: Cambridge, forthcoming). For this reason, this paper contains no discussion of chapter 7 of Noë, where this kind of argument occurs.

[6] Adams and Aizawa, "The Bounds of Cognition," *Philosophical Psychology*, xiv (2001): 43–64, *The Bounds of Cognition*, and "Why the Mind Is Still in the Head."

[7] Rupert, "Representation in Extended Cognitive Systems: Does the Scaffolding of Language Extend the Mind?" in Richard Menary, ed., *The Extended Mind* (Aldershot, Hants: Ashgate, forthcoming), and "Nativism and Empiricism," in Philip Robbins and Murat Aydede, eds., *Cambridge Handbook of Situated Cognition* (New York: Cambridge, forthcoming).

[8] Block, "Review of Alva Noë: *Action in Perception*," this JOURNAL, CII, 5 (May 2005): 259–72.

[9] Prinz, "Putting the Brakes on Enactive Perception," *Psyche*, XII (2006): 1–19.

(COH) based on congenital cataracts, distorting lenses, and retinal stabilization Nor does Block's review discuss the evidence against (COH) based on studies of neuromuscular blockade Prinz's review also involves drawing attention to a causation-constitution distinction and challenges Noë's interpretation of the distorting lenses and retinal stabilization results in ways complementary to that presented here

II REFINING (CAH) AND (COH)

Noë's theory of enactive perception admits of different interpretations and embraces a number of hypotheses, but there is excellent textual evidence for thinking that Noë endorses (COH) [10]

> The central claim of what I call *the enactive approach* is that our ability to perceive not only depends on, but is constituted by, our possession of this sort of sensorimotor knowledge (*op cit*, p 2)
>
> What perception is, however, is not a process in the brain, but a kind of skillful activity on the part of the animal as a whole (*op cit*, p 2)
>
> For mere sensory stimulation to constitute perceptual experience—that is, for it to have genuine world-presenting content—the perceiver must possess and make use of *sensorimotor knowledge* (*op cit*, p 10)
>
> The basic claim of the enactive approach is that the perceiver's ability to perceive is constituted (in part) by sensorimotor knowledge (i e , by practical grasp of the way sensory stimulation varies as the perceiver moves) (*op cit*, p 12)
>
> Most recent work on the relation of perception and action stops short of making the constitutive claim that defines the enactive standpoint (*op cit*, p 18) [11]

Notice that, in the first quotation, it is reasonable to read Noë as acknowledging the difference between (COH) and (CAH) and explicitly endorsing (COH) over (CAH) [12]

Four features of (CAH) and (COH) bear comment First, in investigating these hypotheses, it is common ground that perceptual experience is, in some sense, other than mere peripheral stimulation of the sense organs and something more than sensation. Perceiving is more than a "blooming, buzzing confusion." Perceiving might, thus, involve computational processes generating conceptual representations based on sensory, and perhaps motoric or higher-level cognitive, inputs.

---

[10] Both Block and Prinz allude to some latitude in the interpretation of the view Noë wishes to defend

[11] Cf , Noë, p 215

[12] The same apparently holds for Wilson with regard to the more general versions of (CAH) and (COH) "My point here is not the trivial one that enriched environments can causally produce smarter critters, rather, it is that what metaphysically determines the smartness of at least some critters is their being part of wide cognitive systems" (p 195)

Second, it is probably common ground that in asking whether perceptual experiences are constituted, in part, by the exercise of sensorimotor skills we are asking an important empirical question.[13] Thus, to make this an interesting empirical question, rather than a matter of stipulation, we will want some characterization of "perceptual experience" that is conceptually independent of the exercise of sensorimotor skills. In other words, it would be uninteresting to find that perceptual experiences are constituted, in part, by the exercise of sensorimotor skills if one simply means by "perceptual experiences" those experiences that are constituted, in part, by the exercise of sensorimotor skills. That is, we presuppose that it is conceptually possible to have perceptual experiences that are not so constituted, even if as a matter of empirical fact this is impossible on the grounds that perceptual experiences are so constituted.[14]

The third feature to note concerns what a sensorimotor skill is. To begin with, a sensorimotor skill is a capacity to perform in a certain way, rather than merely possessing some theoretical understanding. It is one thing to know what a poker face is, to have a theoretical understanding of what it is. It is another to have the skill to keep a poker face during a poker hand. Noë also believes that perception involves a kind of theoretical understanding, that perception involves the application of concepts. Thus, he also thinks that perception is, in part, constituted by theoretical understanding. This dimension of Noë's view, however, is not germane to present concerns, since it is a component of Noë's view that is independent of the embodiment of perception and theories of extended cognition If theoretical knowledge resides wholly in the brain, and Noe gives us no reason to think otherwise, then perception's theoretical component does not require either embodiment in or extension into muscles or bones So, in the interests of attending to this big picture, this component of Noë's view will be set aside

---

[13] Cf., Noë, p. 210: "Whether or not neural activity is alone sufficient for an experience is, or at least ought to be regarded as, an empirical question. On what empirical basis, let us ask, is the issue taken to have been settled?" Noë's question here is not identical to the question about whether perception constitutively requires sensorimotor skills, but they are similar in spirit. So, there is little room for Noë to maintain that, while the sufficiency question is empirical, the question about constitution is not.

[14] This assumption is not entirely unproblematic, since in Kevin J. O'Regan and Noë, "What It Is Like to See: A Sensorimotor Theory of Perceptual Experience," *Synthese*, cxxix (2001): 79–103, we find the claim, "We suggest that *perception* could be considered to be the exercise of mastery of this kind of object-related sensorimotor contingency" (p. 88).

Fourth, it should be emphasized that this paper is concerned with synchronic, rather than diachronic claims More specifically, the issue is how, at a given time, perception involves sensorimotor knowledge, if at all Does perceptual experience at a given time either causally or constitutively depend on the exercise of sensorimotor skills? It is a separate matter whether or not the perceptual abilities of any organism must, as a matter of evolutionary fact, require the exercise of sensorimotor skills Perhaps an organism could evolve perceptual capacities in the absence of sensorimotor skills, perhaps not This evolutionary issue, however, is not the present concern It is also a separate matter whether the (normal or proper) development of an individual organism's perceptual abilities depends on the exercise of sensorimotor skills There is experimental evidence to suggest that it does, but that is not the issue here

III EVALUATING THE CASE FOR (COH) VERSUS (CAH)

Why should cognitive scientists suppose that the exercise of sensorimotor mechanisms or skills, in part, constitutes perceptual experiences? Noe offers three types of evidence in support of (COH) the results of removing congenital cataracts, the effects of distorting goggles, and the fading of retinally stabilized images Each of these, however, fails to support (COH) over (CAH)

III 1 Cases of Congenital Cataracts The first source of evidence Noë cites involves the consequences of removing congenital cataracts in adult patients Consider two of the reports to which Noë draws our attention One is Richard L Gregory and Jean G Wallace's description[15] of patient, S B

> S B 's first visual experience, when the bandages were removed, was of the surgeon's face He described the experience as follows He heard a voice coming from in front of him and to one side he turned to the source of the sound and saw a "blur " He realized that this must be a face Upon careful questioning, he seemed to think that he would not have known that this was a face if he had not previously heard the voice and known that voices came from faces (*ibid*, p 366)

---

[15] Gregory and Wallace, *Recovery from Early Blindness A Case Study* (Cambridge Heffer and Sons, 1963), cited by Noë, p 5

Oliver Sacks[16] provides another example

> Virgil told me later that in this first moment he had no idea what he was seeing There was light, there was movement, there was color, all mixed up, all meaningless, a blur Then out of the blur came a voice that said, "Well?" Then, and only then, he said, did he finally realize that this chaos of light and shadow was a face—and, indeed, the face of his surgeon (*ibid.*, p 114)

Clearly, immediately after surgery, the adult patients continue to have some sort of visual perceptual deficit

The existence of this deficit does not, of course, tell us about its nature The cataract patients have sensations, but there still appears to be something missing What? One possibility, the one Noë favors, is that the deficit arises because these patients have not integrated their sensations with their sensorimotor skills The patients are what Noe describes as "experientially blind" (*op cit*, p 4) Another possibility is that growing up with cataracts in place leads to a deficit in sensory processing mechanisms Sensory processes are distorted, hence so are the perceptual processes based on them These possibilities are, of course, not mutually exclusive Growing up with cataracts may well cause both deficits in the production of sensations and in the integration of sensations with the normal sensorimotor apparatus Indeed, this suggests yet another possibility Improper development of either sensation or improper integration of sensation and motor skills could lead to deficits in the sensorimotor apparatus itself

Unless Noe can rule out sensory deficits as the explanation for the behavior of the cataract patients, his argument for (COH) collapses Yet, after providing an articulate statement of this problem, Noe offers the following comments "This objection has some force In section 1 3 I turn to an example of putative experiential blindness [that of distorting lenses] that is not vulnerable to this criticism Taken together the two examples make a strong case for experiential blindness and so for (COH)" (*op cit*, p 7) While conceding some force to the argument in one sentence, two sentences later he appears to think the case of congenital cataracts helps make for a "strong case" for experiential blindness Yet, if the evidence does not show that there is some deficit of sensorimotor integration, it is unclear how it can support (COH) Generally, data that accords with either of two interpretations is not taken to support one or the other The situation here is worse for Noë than just the logic of the situation indicates It is not just that sensory processing deficits are logically possible in cases of congenital cataracts There is, in fact, experimental evidence indicating that congenital cataracts lead to deficits in sensory processing Perhaps there are sensorimotor integration problems and sensorimotor problems, but there is evidence for sensory deficits

---

[16] Sacks, *An Anthropologist on Mars Seven Paradoxical Tales* (New York Knopf, 1995), cited by Noë, p 5

Several recent studies have indicated that even after surgical removal of dense congenital cataracts, patients suffer deficits in grating acuity, spatial contrast sensitivity, temporal contrast sensitivity, peripheral vision, stereo acuity, perception of global form, and perception of global motion [17] While many of these studies appeared too recently to have been incorporated into Noë's book, there is another source of evidence bearing on the nature of the deficits in patients with congenital cataracts For over forty years, neuroscientists have been investigating the neuroanatomical and neurophysiological effects of various forms of visual sensory deprivation in nonhuman animals Among the best known are classic early experiments conducted by Torsten Wiesel and David Hubel [18] Most relevant for present purposes are studies involving binocular lid suture for the first several weeks after an animal's birth. These interventions, like large dense congenital cataracts, allow the projection of diffuse light onto the retina, while blocking the highest frequency spatial patterns in the stimuli. That is, eyelid sutures and large dense congenital cataracts allow only broad patterns of light and dark, rather than fine details, to play upon the retina. Even this restricted type of sensory deprivation has a substantial literature. In a 1982 review, Murray Sherman and Peter Spear reported that binocular lid suture in cats during the critical or sensitive period produces any number of abnormalities in striate cortex.[19] After the sutures are removed, deprived striate cortex is less responsive to visual stimulation than is normal striate cortex. In deprived cats, fewer cells respond to standard stimuli and even those cells that do respond have lower peak response rates. In addition, among cells that remain responsive, a large proportion have abnormal receptive fields. In normal cats, a large fraction of cortical cells have excitatory and inhibitory regions that make them sensitive to the orientations of edges, bars, and lines. In bilaterally lid sutured cats, however, the inhibitory portions of the receptive fields of these cortical cells are weak or absent. This will likely make the perceptual experience of borders faint or fuzzy, suggesting a sensory explanation of some of the perceptual deficits found in Gregory and Wallace's patient, S. B., and Sacks's patient Virgil. Further, in normal cats, many of the cells in striate cortex are sensitive to the direction of motion of edges, bars, and lines. In bilaterally lid sutured cats, many of the responsive cortical cells lose their sensitivity to the direction of motion. This might explain any failure of sensorimotor coordination found in patients after bilateral congenital cataract removal. Finally, bilaterally lid sutured cats also lose sensitivity to binocular disparity, thereby losing an important depth cue. This purely sensory deficit might explain the difficulty patients might have in negotiating the three-dimensional world, hence obviating the need to hypothesize experiential blindness as the cause of a patient's behavioral deficits.

---

[17] Eileen E Birch, David Stager, Joel Leffler, and David Weakley, "Early Treatment of Congenital Unilateral Cataract Minimizes Unequal Competition," *Investigative Ophthalmology and Visual Science*, XXXIX (1998) 1560–66, Elizabeth R Bowering, Daphne Maurer, Terri L Lewis, and Henry P Brent, "Sensitivity in the Nasal and Temporal Hemifields in Children Treated for Cataract," *Investigative Ophthalmology and Visual Science*, XXXIV (1993) 3501–09, Lewis, Maurer, and Brent, "The Development of Grating Acuity in Children Treated for Unilateral or Bilateral Congenital Cataract," *Investigative Ophthalmology and Visual Science*, XXXVI (1995) 2080–95, David Ellemberg, Lewis, Maurer, and Brent, "Influence of Monocular Deprivation During Infancy on the Later Development of Spatial and Temporal Vision," *Vision Research*, XL (2000) 3283–95, Ellemberg, Maurer, Sonia Brar, and Brent, "Better Perception of Global Motion after Monocular than after Binocular Deprivation," *Vision Research*, LVII (2002) 169–79, Ellemberg, Lewis, Maurer, Chang Hong Liu, and Brent, "Spatial and Temporal Vision in Patients Treated for Bilateral Congenital Cataracts," *Vision Research*, XXXIX (1999) 3480–89, Lewis, Ellemberg, Maurer, Fran Wilkinson, Hugh R. Wilson, Melanie Dirks, and Brent, "Sensitivity to Global Form in Glass Patterns after Early Visual Deprivation in Humans," *Vision Research*, LXII (2002) 939–48, L Mioche and M T Perenin, "Central and Peripheral Residual Vision in Humans with Bilateral Deprivation Amblyopia," *Experimental Brain Research*, LXII (1986) 259–72, Milan E Tytla, Lewis, Maurer, and Brent, "Stereopsis after Congenital Cataract," *Investigative Ophthalmology and Visual Science*, XXXIV (1993) 1767–73, Milan E Tytla, Maurer, Lewis, and Brent, "Contrast Sensitivity in Children Treated for Congenital Cataract," *Clinical Vision Sciences*, II (1988) 251–64

[18] Hubel and Wiesel, "Receptive Fields of Cells in Striate Cortex of Very Young, Visually Inexperienced Kittens," *Journal of Neurophysiology*, XXVI (1963) 994–1002, and "Comparison of Effects of Unilateral and Bilateral Eye Closure on Cortical Unit Responses in Kittens," *Journal of Neurophysiology*, XXVIII (1965) 1029–40

The point is not that the current clinical and experimental results definitively establish a purely sensory explanation for patient deficits. There are clear differences between the human cases and the animal experiments. For example, in the animal experiments, the eyelids are sutured before the animal's eyes open. In human patients, the opacification of cataracts takes place over several weeks and months, thereby allowing some amount of visual experience. Further, the cataracts need not cover the entire pupil, thereby allowing "slits" of relatively normal stimulation around the edges of the cataract. Still further, the presence of cataracts in humans can induce behavioral abnormalities, such as squinting, that are not found in the animal.[20] These differences notwithstanding, the point remains that the possibility of congenital cataracts giving rise to sensory deficits is not a mere possibility, it is an empirical possibility supported by a serious body of evidence that challenges Noë's interpretation of the data.

---

[19] Sherman and Spear, "Organization of Visual Pathways in Normal and Visually Deprived Cats," *Physiological Reviews*, LXII (1982): 738–855.

The foregoing empirical problems are just the beginning of the difficulties in Noe's reliance on observations of post-operative cataract patients to argue for (COH). Suppose, for the sake of argument, that in fact these patients are experientially blind in Noë's sense. That is, suppose that they do not or cannot integrate their sensations with patterns of movement and thought. Does this suffice to establish (COH)? No. There still remain two ways in which this conclusion falls short of supporting (COH). Noe's conception of experiential blindness is a kind of perceptual deficit that he sets alongside sensory blindness, it is an inability to integrate sensations with patterns of movement and thought. But, there are, of course, many kinds of perceptual blindness. There is achromotopsia (colorblindness) and this of several types. There is akinetopsia, motion blindness. There is face agnosia and form agnosia. A person who is color blind is not entirely lacking in perceptual capacities, only the capacity to discriminate some colors. A person who is motion blind is not completely blind, only blind to motion. Similarly for face and form agnosia. And, indeed, the same for Noe's experiential blindness. As a conceptual possibility, it appears to be possible that some humans might perceive things, only without these perceptions being integrated into patterns of personal movement and thought. The only thing that might preclude this conceptual possibility is if Noe's concept of perceptual experience simply includes being constituted, in part, by sensorimotor skills. But, then, Noë's claim that perceptual experience is constituted, in part, by sensorimotor skills just becomes some sort of logical truth, rather than an empirical hypothesis. The examples of cataract patient are then irrelevant.

---

[20] Each of these possibilities is described in a case study in Ione Fine, Harvey S. Smallman, Peter Doyle, and Donald I. A. Macleod, "Visual Function before and after the Removal of Bilateral Congenital Cataracts in Adulthood," *Vision Research*, LXII (2002) 191–201.

Even if we concede that the cataract patients are experientially blind, there still remains one more problem in using this to support (COH) One must ask what is responsible for the patients's failure to integrate sensory stimulation with sensorimotor skills What explains this failure? Noe concludes that the failure is due to the loss of a constituent of the perceptual process, namely, the exercise of sensorimotor skills No exercise of sensorimotor skills, no perceptual experience That is, he thinks that (COH) explains the failure of integration, he thinks that (COH) explains the experiential blindness But, (CAH) seems to offer just as good an explanation of the putative experiential blindness as does (COH) According to (CAH), the reason post-operative cataract patients are experientially blind is that their sensorimotor skills are not yet causally connected to their sensory apparatus in the proper way It is not that the exercise of sensorimotor skills are essential to perception because they are constitutive, in part, of perception, it is that these sensorimotor skills are essential for normal perception because they have an essential causal role in shaping perceptions and the normal causal connections have been disrupted

So, it appears that Noe's appeal to post-operative congenital cataract patients fails to establish (COH) in three ways First, the cataract patients may well suffer from sensory deficits of one or more types which could entirely account for any perceptual or sensorimotor deficits Second, even if cataract patients are experientially blind in Noe's sense, this still fails to establish (COH) for two reasons First, these patients might be experientially blind, but this does not imply that they must be totally blind as (COH) requires Second, (COH) provides no better explanation of experiential blindness than does (CAH)

III 2 Distorting Lenses Since the late nineteenth-century experiments by George Malcolm Stratton, psychologists have been interested in the effects of long-term use of distorting lenses of one sort or another [21] Noe wishes to appeal to them in support of (COH) Noe is correct in his claim that these experiments provide cases of experiential blindness that overcome the problem of possible sensory deficits in cataract patients Nevertheless, he fails to recognize the gap between the discovery of experiential blindness and the confirmation of (COH) that we just rehearsed above Consider, then, Ivo Kohler's description[22] of the effects of wearing spherical prism lenses,

---

[21] George Malcolm Stratton, "Some Preliminary Experiments on Vision without Inversion of the Retinal Image," *Psychological Review*, III (1896) 611–17, "Upright Vision and the Retinal Image," *Psychological Review*, IV (1897) 182–87, and "Vision without Inversion of the Retinal Image," *Psychological Review*, IV (1897) 341–60 and 463–81

> During visual fixations, every movement of my head gives rise to the most unexpected and peculiar transformations of objects in the visual field The most familiar forms seem to dissolve and reintegrate in ways never before seen At times, parts of figures run together, the spaces between disappearing from view, at other times, they run apart, as if intent on deceiving the observer Countless times I was fooled by these extreme distortions and taken by surprise when a wall, for instance, suddenly appeared to slant down the road, when a truck I was following with my eyes started to bend, when the road began to arch like a wave, when houses and trees seems to topple down, and so forth I felt as if I were living in a topsy-turvy world of houses crashing down on you, of heaving roads, and of jellylike people (*ibid*, p 64)

What appears to be going on here is a failure to integrate sensations into sensorimotor skills, that is, we appear to have experiential blindness This is the point of Noe's example

Nevertheless, as we saw in the previous section, the existence of experiential blindness does not suffice to establish (COH) Recall that experiential blindness, as Noë defines it, is just a particular kind of perceptual deficit in just the way achromotopsia and akinetopsia are But, one can have this deficit without being entirely blind In the case of the subject, K, even Noe admits that "K is not completely blind, to be sure, he recognizes the trucks, the trees, and so forth But nor is he completely able to see His visual world is distorted, made unpredictable, and topsy-turvy" (*op cit*, p 8, cf p 9) If K is able to synthesize the sensations caused by trucks, trees, people and so forth, into forms that enable him to recognize them as trucks, trees, and people, then there seems to be some reasonable basis for saying that he perceives the trucks, trees, people, and so forth Thus, K perceives things, even though experientially blind K perceives things despite the misalignment between his vision and his sensorimotor skills So, Noë does not give us a case where there is no exercise of sensorimotor skills, hence no perceptual experience Noë does not give us a case of experiential blindness bringing on an absence of perceptual experience So, the observations regarding K do not support (COH)

Noë is evidently aware of the force of this objection and subsequently tries to blunt it Noë observes that the left-right distorting lenses only affect one's perceptions involving motions to the left or right, not those involving motions up and down[23] Indeed, the left-right goggles do not influence perceptions of colors or light and dark Viewed abstractly, what Noe is saying is that failing to exercise sensorimotor skills of type α leads to loss of perceptual experiences of type β But, this could, at most, be used to support (COH*), namely, that perceptual experiences of type β are constituted, in part, by the exercise of sensorimotor skills of type α This is obviously weaker than (COH) So, even if (COH*) were true, this would not suffice to establish (COH) Further, (COH*) runs a greater risk than does (COH) of degenerating into an uninteresting stipulative definition We need some characterization of perceptual experiences of type β that does not stipulate that they are simply those abilities constituted, in part, by the exercise of sensorimotor skills of type α

---

[22] Kohler, "Formation and Transformation of the Perceptual World," *Psychological Issues*, III (1964) 1–173, cited by Noë, p 8

Both (COH*) and (COH) face a further problem, namely, they do not appear to provide a better explanation of experiential blindness than would hypotheses (CAH*) or (CAH) This last point will be sketched for (COH) As in our discussion of the cataract patients, we can ask for an explanation of the experiential blindness of those who wear distorting goggles Why is it that those who wear distorting lenses have their sensations decoupled from their sensorimotor apparatus? Noe proposes (COH), that a constitutive element of perceptual experiences is lost, namely, that there is no exercise of sensorimotor skills But, (CAH) seems to offer as good an explanation of experiential blindness as does (COH) (CAH) tacitly supposes that certain visual perceptions are influenced by two types of factors, the input from the retina and input from the sensorimotor mechanisms, such as those in the vestibular system, that compensate for bodily motions The distorting lenses, however, disrupt the relationship that normally exists between the retinal input and the motion compensating mechanisms Thus, the lenses distort the coordination of retinal signals and bodily signals There is no need to adopt (COH) for this explanation In view of this last point, the observations on the effects of distorting lenses provide no reason to think that any perceptual experiences are in any way constituted by the exercise of any sensorimotor skills

III 3 Visual Fading of Retinally Stabilized Images For over fifty years, it has been known that images projected to a fixed location on the

---

[23] It would presumably be somewhat harder for Noë to run this line in K's case, since he wore spherically distorting lenses, rather than left-right or up-down reversing lenses Perhaps he can maintain that the distorting lenses do not eliminate the sensorimotor skills of eye movements, hence that perception can persist But, this would evidently undercut the appeal to lenses, since they would no longer be cases where there is no sensorimotor skill, hence no perception

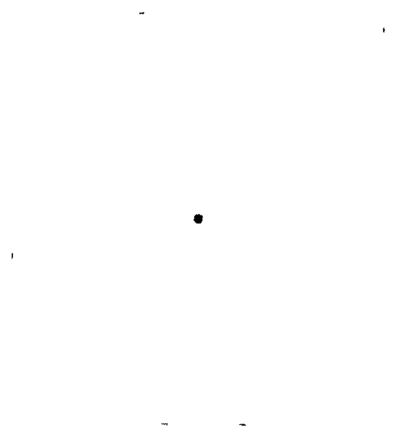


Figure 1

retina fade due to sensory adaptation.[24] Rods and cones cease to respond to a uniform pattern of illumination. While the use of laboratory equipment is most effective for stabilizing images on the retina, one can get a sense of this kind of adaptation by fixating on the black spot in the center of Figure 1. With a steady fixation, the gray annulus will fade from view, but will return with slight eye movements.

Although Noë claims that visual fading "suggests that some minimal amount of eye and body movement is necessary for perceptual sensation" (*op. cit.*, p. 13), this suggestion does nothing to support (COH). In the first place, this provides no grounds for extrapolating from a putative need for motion in vision to a need for motion in other sense modalities, such as olfaction. Second, it is simply not true that some minimal amount of eye or body movement is necessary even for visual perception. Even if the eye and body are entirely stationary, the fading can be prevented simply by moving images across the retina or changing the pattern of light projected onto the retina. The reader may have someone hold Figure 1 steady before her until the annulus fades, then have that person move the figure. The annulus will then reappear. It is the change of stimulation, not necessarily the eye or body motions, or sensorimotor skills, *per se* that are relevant. Third, even if eye or body motions were necessary to prevent visual fading, that would not show that the exercise of sensorimotor skills is necessary to prevent visual fading Parkinsonian tremors, for example, might induce sufficiently large motions of the eye and body to prevent visual fading, but the tremors are not the exercise of sensorimotor skills One cannot simply identify eye and bodily motions with the exercise of sensorimotor skills Fourth, suppose, contrary to fact and just for the sake of argument, that the exercise of sensorimotor skills is necessary for visual perception Even that would not show that exercising certain sensorimotor skills partially constitutes certain visual perceptual experiences, rather than that the exercise of those skills is causally necessary for visual experiences (CAH) is not the view that the exercise of sensorimotor skills is causally necessary for perceptual experiences As we shall see below, experiments with neuromuscular blockade suggest otherwise (CAH) only maintains that the exercise of sensorimotor skills makes a causal contribution to the perceptual experiences one has If, however, it turns out, contrary to fact, that it is only by the exercise of certain sensorimotor skills that one has certain perceptual experiences, (CAH) can explain this by appeal to the idea that the exercise of certain sensorimotor skills happens to be the only causal mechanism for producing those experiences Thus, (CAH) can provide as good an explanation of the necessity of deployed sensorimotor skills for perceptual experiences as does (COH) So, retinal fading provides no evidence that any perceptual abilities are constituted, in part, by the use of sensorimotor skills There really is a surprising gap between the actual empirical data and the interpretation Noe wishes to place on them

---

[24] L. A. Riggs and F. Ratliff, "The Effects of Counteracting the Normal Movements of the Eye," *Journal of the Optical Society of America*, XLII (1952): 872–73, and R. W. Ditchburn and B. L. Ginsborg, "Vision with a Stabilized Retinal Image," *Nature*, CLXX (1952): 36–37 are sometimes cited as classic papers on the topic.

IV THE CASE AGAINST (COH) PARALYSIS

Given that the evidence Noe presents fails to support (COH) over (CAH), one might suspect that this is a consequence of an insignificant difference between the two, that (COH) makes no empirical predictions that (CAH) does not Yet, this is incorrect In fact, (COH) makes some predictions that are disconfirmed These are cases in which paralysis prevents the exercise of sensorimotor skills, but where perceptual experiences continue Noe hints at two strategies that one might use to address this problem of paralysis The first is by appeal to sensorimotor knowledge as theoretical and the other by appeal to the incompleteness of paralysis The first approach does nothing to save the embodiment of perception, the second only works for a limited range of types of paralysis

Noë writes apparently interchangeably about sensorimotor skills and sensorimotor knowledge This is unproblematic, provided the assumption that skills in the muscles and peripheral nerves constitute a kind of knowledge But, Noe also comments on cases of motor paralysis and optic ataxia (the inability to coordinate motor actions with sensations) in ways that suggest an alternative conception of sensorimotor knowledge as a species of theoretical knowledge.

> The existence of optic ataxia, therefore, does not undercut the enactive view, for from the fact that a patient suffers optic ataxia, it doesn't follow that he or she lacks the relevant sensorimotor knowledge. What would undercut the enactive approach would be the existence of perception in the absence of bodily skills and sensorimotor knowledge which, on the enactive view, are constitutive of the ability to perceive. Could there be an entirely inactive, *inert* perceiver? (*op. cit.*, p. 12)

What Noë suggests in the second sentence is that perceptual abilities must constitutively involve either bodily skills or (theoretical) sensorimotor knowledge. In optic ataxia one loses the bodily skills but retains the theoretical sensorimotor knowledge. Therefore, the enactive view predicts that those suffering from optic ataxia should not be completely without perception. This reading is, however, somewhat strained by the concluding question. An entirely inactive, inert perceiver would be one that is immobilized or paralyzed, not one necessarily lacking theoretical sensorimotor knowledge. So, in this passage it still looks as though Noë is committed to the exercise of bodily skills in muscles and peripheral nerves being constitutive of perceptual experience. But, then again, just a paragraph later, Noë writes the following,

> More important, paralysis does not undermine the paralyzed person's practical understanding of the ways movement and sensory stimulation depend on each other. Even the paralyzed, whose range of movement is restricted, understand, implicitly and practically, the significance of movement for stimulation. They understand, no less than those who are not disabled, that movement of the eyes to the left produces rightward movement across the visual field, and so forth. Paralyzed people can't do as much as people who are not paralyzed, but they can do a great deal, whatever the scope of their limitations, they draw on a wealth of sensorimotor skill that informs and enables them to perceive. (*op. cit.*)

Again, the initial sentences of this passage could be taken to mean that theoretical knowledge of the relations between bodily motions are sufficient for preserving perception in paralyzed individuals. But, of course, there are the closing comments on what paralyzed people can and cannot do, which at least suggests that Noe is sticking by the idea that perception requires the exercise of some bodily sensorimotor skills.

Noë can, of course, define the theory of enactive perception as he wishes, either allowing theoretical knowledge to preserve perception or not. Be Noë's view what it may, an appeal to theoretical knowledge will not support the embodiment of perception in nerves and periph-

KP 2686

eral muscles Insofar as theoretical knowledge is constituted entirely by what is in the brain, showing that perceptual experience is constituted, in part, by the exercise of theoretical knowledge does not suffice to show that perceptual experience is constituted, in part, by the exercise of sensorimotor skills and processes in muscles and peripheral nerves So, insofar as sensorimotor knowledge is taken to require the active use of muscles and peripheral nerves, it appears that the enactive theory is committed to there being a total loss of perception in cases of total paralysis

Noe's second strategy in dealing with paralysis, therefore, is to note that in the case of quadriplegics, paralysis is not complete Here is the relevant text

> Paralysis is certainly not a form of blindness But isn't that precisely what the enactive theory requires, that the paralyzed be experientially blind? No The enactive theory requires that perceivers possess a range of sensorimotor skills It seems clear that quadriplegics have the pertinent skills Quadriplegics can move their eyes and head, and to some extent, at least with help from technology, they can move their bodies with respect to the environment (e g , by using a wheelchair) (*op cit*, p 12)

Noë is apparently equating paralysis with quadriplegia Quadriplegia, however, is not the only source of paralysis In particular, there are forms of paralysis that arise from interference with the neuromuscular junction

There is a broad class of neuromuscular blockers that inhibit neuromuscular function by binding to the muscle receptors for the neuromuscular transmitter acetyl choline Acetyl choline cannot cause normal neuromuscular contraction, since the muscle relaxer blocks its binding sites on muscles For decades, the use of neuromuscular blockade has widely been recognized to induce paralysis, but not loss of consciousness Muscle relaxers are, thus, quite different from general anesthetics Muscle relaxers have long been an important tool in the anesthesiologist's toolkit, enabling the surgeon to better control the body position of patients They are also exceptionally useful in suppressing involuntary motor responses, such as the gagging and coughing that occur during intubation

These facts bear on the enactive theory of perception since patients are frequently aware of sounds, voices, and objects during surgery that combines both general anesthetic and neuromuscular blockade[25] In an anonymous 1979 editorial in the *British Journal of*

---

[25] In a forthcoming book, *The Bounds of Cognition*, Adams and Aizawa also review complementary studies of the effects neuromuscular blockade on perception

Anaesthesia, a woman with medical experience reported in graphic detail her recollections of being aware during a Caesarian section under general anesthesia and neuromuscular blockade [26] Although immobilized by the anesthesiologist, she remembers feeling the pain of the incision, hearing her baby crying, and feeling the insertion of a nasogastric tube In a more recent case study, a 74-year old woman recalled that during her operation "1) she felt pain during the incision of the abdomen, 2) she heard the operator say, 'It is difficult to remove all tumors because the adhesion is very strong,' and 3) she remembered someone had been walking around her "[27] These post-operative recollections were independently confirmed by the surgical staff Because of the obvious importance of this phenomenon, there have been a number of studies of its frequency in hospitals in the United States and Europe In one study involving 45 patients, all recalled hearing sounds or voices, 33 understood and remembered conversations, 21 had visual perceptions, 29 felt being touched, 6 recalled moderate pain, and 8 recalled severe pain [28] Noe focuses on the role of sensorimotor skills in visual perception, but the cases of awareness of voices, sounds, and pain under anesthesia remind us of the need to consider other perceptual modalities

Paradoxically for the enactive theory, one large-scale prospective study of awareness during surgery found that recollection of awareness is *more* probable when general anesthetics are used in combination with muscle relaxers than when general anesthetics are used alone [29] If the enactive theory is correct, this is surprising since the theory might well be taken to predict that the more extensive the immobilization, the less likely there is to be perception If, however, sensorimotor skills are only a causal influence on what one perceives, then it is not at all surprising that the use of muscle relaxers would increase the incidence of awareness during surgery With the body incapacitated, it is much more difficult for anesthesiologists to detect the fact that a patient is not sufficiently anesthetized When patients are more lightly anesthetized, they are more likely to be aware This makes sense on (CAH)

---

[26] "On Being Aware," *British Journal of Anaesthesia*, LI (1979) 711–12

[27] Shihiro Miura, Satoshi Kashimoto, Toshiaki Yamaguchi, and Takashi Matsukawa, "A Case of Awareness with Sevoflurane and Epidural Anesthesia in Ovarian Tumorectomy," *Journal of Clinical Anesthesia*, XIII (2001) 227–29

[28] D Schwender, H Kunze-Kronawitter, P Dietrich, S Klasing, H Forst, and C Madler, "Conscious Awareness during General Anaesthesia Patients' Perceptions, Emotions, Cognition, and Reactions," *British Journal of Anaesthesia*, LXXX (1998) 133–39

[29] Rolf H Sandin, Gunnar Enlund, Peter Samuelsson, and Claes Lennmarken, "Awareness during Anaesthesia A Prospective Case Study," *The Lancet*, CCCLV (2000) 707–11

V CONCLUSION

The upshot is that the evidence from congenital cataract patients, subjects wearing distorting lenses, and retinal fading fails to support the radical hypothesis that perception is constituted, in part, by the exercise of sensorimotor skills Rather, the evidence is perfectly consistent with what mainstream cognitive psychology has long believed, namely, that perception is causally influenced by sensorimotor skills Further, cases of complete paralysis found in the use of neuromuscular blockade show that perception is possible without the exercise of sensorimotor skills

If the enactive view provides a distorted account of the embodiment of perception, what is a more plausible approach? For the present, only a few all-too-brief comments can be made

To begin with, (CAH) itself needs further refinement and clarification For one thing, it is evidently false to say that all perceptual experiences are causally influenced by the exercise of sensorimotor skills As we saw in the cases of neuromuscular blockade, there are some perceptual experiences that are essentially the same, whether one exercises sensorimotor skills or not One perceives speech, for example, in essentially the same way whether paralyzed by neuromuscular blockade or not It is much more plausible to say that some perceptual experiences are causally influenced by the exercise of sensorimotor skills A second consideration is that we may have to distinguish between the possible roles of motors signals to execute bodily movements which might be entirely within the nervous system and the possible role of the actual flesh and blood, bone and cartilage of the body For example, during a blink, the eye is completely closed for about 1/6th of a second with the amount of light striking the retina reduced by about 90% [30] These blackouts are noticeable, but they are not of the same magnitude as would be noticed were the lights in a room to be dimmed by the same amount for the same length of time One hypothesis that has been advanced to account for this is that the brain signals the eyes to blink and sends an ancillary signal to the visual system that somehow suppresses visual processing during the blink What a person perceives over the span of an eye blink may, therefore, be a function of both a purely internal brain-bound information processing mechanism and the physical movement of the eye lid over the eye A third consideration is that attention to the role of sensorimotor skills in perception is also too narrow a perspective on the embodiment of perception The body influences perception by means other than the execution of sensorimotor skills For example, the external structures of the ears, the pinnas, appear to facilitate the localization of sound by the way in which sound waves bounce around the folds of the ear [31] Finally, there are diachronic considerations about the role of the body in ontogenetic develop of perception and the role of the body in the evolution of perceptual capacities There is clearly a vast amount that remains to be said to do justice to what constitutes the embodiment of perception

---

[30] See Randolph Blake and Robert Sekuler, *Perception* (New York McGraw-Hill, 2002, 5th edition), p 40, for references and further details

KENNETH AIZAWA

Centenary College of Louisiana

---

[31] See Blake and Sekuler, pp 417f , for references and further details

MACROSCOPIC MIXTURES*

This paper takes up issues related to the notion of chemical substances—henceforth, simply *substances*—arising from their mereological and modal features. Related notions are elements and compounds, into which substances are subdivided, and the general notion of mixture, which as a special case might involve several substances, but covers other cases too. These are essentially macroscopic concepts. Some of the chemical arguments for this claim have been presented elsewhere.[1] The present paper is a metaphysical treatment of matter as categorized by the major chemical concepts bearing on distinctions between homogeneous and heterogeneous matter, and between a single and several substances.

Section I introduces some of the central concepts, explaining and illustrating the general notion of mixture, and indicating some connections with issues related to the discussion of natural kinds. Important distinctions between features that are and features that are not time-dependent are emphasized, as well as a corresponding modal distinction between features that do and do not hold necessarily. The argument gets under way in earnest in the section II with a critique of W.V. Quine's analysis of the mass predicate character of substance predicates, which he described in mereological terms. Without questioning the basic principles of extensional mereology, I criticize a certain point of his analysis, which raises two issues. One is connected with a technical matter about extending the mereological analysis of mass predicates to deal with relational mass predicates, and leads me to emphasize the macroscopic character of the concepts under discussion. The second issue concerns how mixtures are to be understood, which is pursued in section III. I introduce a view of mixture that I call the *Stoic view*, which I contrast with and defend against a view I call the *Aristotelian view*. (This will involve an analytic comparison between views which are, as the names suggest, clearly

---

* I would like to thank Robin Hendry, Fred Stoutland, and Michael Weisberg for helpful comments on earlier drafts of this paper.

[1] See Jaap van Brakel, "The Chemistry of Substances and the Philosophy of Mass Terms," *Synthese*, LXIX (1986): 291–324; "The Nature of Chemical Substances," in Nalini Bhushan and Stuart Rosenfeld, eds., *Of Minds and Molecules* (New York: Oxford, 2000), pp. 162–84; and *Philosophy of Chemistry: Between the Manifest and the Scientific Image* (Leuven: University Press, 2000); Paul Needham, "What Is Water?" *Analysis*, LX (2000): 13–21; and "The Discovery That Water Is H_2O," *International Studies in the Philosophy of Science*, XVI (2002): 201–22.

0022-362X/07/0401/26–52

derived from ancient theories. But no attempt is made to justify them as interpretations by detailed examination of the classical texts, and limitations of space preclude any detailed examination of variations on the two central views which the details of the present analysis may well suggest.) The spatial character of mereological relations between quantities is developed here, leading to a distinction between two categories of mass predication, and different conceptions of elements (noncompound substances). The section concludes with some comments about the understanding of cooccupancy from the macroscopic perspective. Modality enters the discussion in section IV with a suggestion for how the notion of potential presence central to the Aristotelian conception of mixture might be understood, and contrasted with modal claims involved in the Stoic view. The rigidity of mereological parthood is introduced here, and necessary preservation or otherwise of properties such as being water iff being H_2O, and being the same substance, is discussed. The final section adds some concluding remarks about potentiality on the two views and summarizes some of the main points.

I

Chemists use the term 'mixture' on the whole more broadly than philosophers have used the term, but in one sense, more narrowly. For chemists, the term covers *solutions* (that is, uniform homogeneous mixtures of several substances) and heterogeneous mixtures (comprising matter in distinct homogeneous parts, which chemists call phases, such as the oily and the watery parts of a liquid mixture, or a vapor phase over a volatile liquid), but not single compounds in a uniform homogeneous state (that is, in a single phase). As used here, 'mixture' does not exclude this last case, and in this respect the term is used as it was before a distinction between compounds and solutions was introduced on the basis of the law of constant proportions.

Mixtures are important in chemistry. Features characterizing substances such as water are not restricted to properties of pure water, nor even quantities of matter comprising mainly water.[2] What, exactly, is meant by matter comprising mostly some particular substance may depend on whether proportions are measured in gravimetric or mole units, and is not determinable antecedently to theoretical commitment to some understanding of what substances are. But on any account of proportions, the important features of substances are not confined to those exhibited in matter where it is the predominant component The whole range of composition is important Consider how composition affects a mixture of water and ether Adding a little water to ether produces a homogeneous solution of water in ether Adding more water leads to the formation of two separate layers, that is, distinct liquid phases, the upper one richer in ether and the lower richer in water Each phase is homogeneous, but the mixture as a whole is heterogeneous The composition of each phase is dependent on the temperature and pressure, but not, in accordance with Gibbs's phase rule, on the masses of either component Accordingly, continuing to add more water at a given temperature and pressure reduces the mass of the upper layer and increases the lower to maintain the concentrations in each layer, until finally the mixture becomes homogeneous again and a single phase remains comprising a solution of ether in water The example illustrates the general case of a mixture comprising several phases over which distinct substances are distributed in proportions governed by prevailing conditions in accordance with laws of macroscopic theory Water at its triple point, when solid, liquid, and gas are at equilibrium, is a special case of a mixture comprising a single substance simultaneously exhibiting several phases Solutions such as liquid brine (salt in water) or air (oxygen, nitrogen, and other substances in the gaseous state, ignoring dust) are special cases of several substances in a single phase

---

[2] Jessica Brown, "Natural Kind Terms and Recognitional Capacities," *Mind*, CVII (1998): 275–303, and Barbara Abbot, "Water = H_2O," *Mind*, CVIII (1999): 145–48.

Under the conditions which allow for ordinary chemical phenomena, then, matter occurs in the solid, liquid, or gas phases, and sometimes in several solid and liquid phases Distinctions of phase are just as important to the understanding of mixtures as are distinctions of substance, and the corresponding predicates have analogous mereological properties which will be discussed in connection with the two views of mixture Even single substances exhibit different phases Hilary Putnam's discussion of the property of being the same liquid as this sample of water is misleading in this respect in so far as it suggests that being the same substance is restricted to a particular phase, namely the property of being liquid It might be Some substance terms do carry a phase restriction, as do 'quartz' (a particular crystalline form of silicon dioxide) and 'rhombic sulphur' (a particular crystalline form of sulphur) But this is not how substances are understood on the general view of mixture outlined above (or when water is said to have a triple point at 0 01 °C and 4 58 mm, Hg—a property ascribed not to liquid water, but to water) [3] Moreover, the usages should not be confused. If 'water' were understood as Putnam evidently understands the term, to refer exclusively to the liquid, then although 'Water is H_2O' is true the converse, 'H_2O is water' (in the mass quantification sense 'All H_2O is water'), would not be. In the general phase-independent sense of the term, however, an equivalence does hold between the terms 'is water' and 'is H_2O' because they are true of exactly the same things.

---

[3] It is not objected to Lavoisier's method of determining the composition of water, by boiling the liquid and reducing the steam over heated charcoal, that in decomposing

Strengthening the equivalence to a necessary equivalence provides a criterion for substitutivity of the predicates for one another in modal contexts. This is not a claim about rigidity. Rigidity is defined in quantified modal logic in terms of identity by the two conditions

$$\text{(R) (i) } x = y \supset \Box(x = y) \quad \text{and (ii) } x \neq y \supset \Box(x \neq y),$$

and that is how the term is used here. Predicates cannot meaningfully flank the identity sign, and because substance terms are predicates, it is not meaningful to describe them as rigid. The same goes for complex predicates formed from substance predicates like 'is water iff is H_2O', which are therefore not rigid designators. This does not automatically dissolve the underlying issue since suitable analogues can be formulated. The necessary preservation of being water iff being H_2O yields an analogue of (Ri) which is true, although not a modal theorem.

Neither being the same substance nor its negation, as understood here, is necessarily preserved. The predicate 'is water iff is H_2O' does not express the relation of being the same substance. If this is not already apparent from the syntax of the terms, it should be once it is clear that although this predicate is true, indeed necessarily true, of all matter, being the same substance is certainly not. Quantities of water and aluminium oxide are not the same substance. Neither is a quantity of water *always*, and certainly not necessarily, the same substance. Further, some of what is not water may, and often does, become water. Water, unlike helium, neon, and argon, is a typical chemical substance, participating in innumerable chemical reactions in which it is either created or destroyed. The same substance relation involves temporal qualification, which will be accommodated by counting times among its relata. The other relata are the quantities of matter to which substance and phase properties apply.

---

the gas it is not water that is decomposed. Indeed, what sense would it make to say that water is composed of hydrogen gas and oxygen gas? None of the temperature ranges, at any given pressure, over which these three substances are in the solid, liquid, and gas phases coincide.

Calcutta University Library
ALIPORE CAMPUS

Quantities are understood in a way which avoids the suggestion of paradox in the formulation of chemical change involving the creation and destruction of matter Taken literally, this latter notion is certainly foreign to any modern conception of chemical processes because of Lavoisier's principle of the constancy of mass throughout a chemical reaction, which is taken to indicate that matter is neither created nor destroyed Any appearance of paradox is straightforwardly dispelled by clearly distinguishing the subject of predication from the predicate expressing a substance property quantities of matter can bear substance properties at one time and perhaps not at another, without gain or loss of parts The macroscopic character of quantities, which occupy regions of space for intervals of time, is developed in the course of the paper Quantities will be referred to by singular terms which can meaningfully flank the identity sign, and which will be treated along the lines of Kripke's interpretation of variables in quantified modal logic They will be treated as mereologically rigid, which implies that they satisfy the rigidity conditions

II

Substance predicates are apparently mass predicates, applying to all of what they apply to, a quantity which is water is all water Two general features have been attributed to mass predicates involving mereological concepts The *distributive condition* applied to water says that any part of what is water is water It is to be found in Aristotle, who says that when the mixing of ingredients results in the formation of a compound, "the compound *must* be uniform—any part of such a compound is the same as the whole, just as any part of water is water "[4] Using variables π, ρ, σ, to range over quantities of matter[5] and '$\subseteq$' for the mereological part relation,[6] a predicate $\varphi(\pi)$ is distributive iff

$$\varphi(\pi) \wedge \rho \subseteq \pi \supset \varphi(\rho) \tag{1}$$

In recent philosophy, the distributive condition has been supplemented with a *cumulative condition* Quine says that "mass terms like 'water', 'footwear', and 'red' have the semantical property of referring cumulatively any sum of parts which are water is water ”[7] Where ‘$\cup$’ is the dyadic operation of mereological sum, a predicate, φ, is cumulative if it satisfies

---

[4] *DG* I 10, 328^{a}10f All quotations from Aristotle are taken from *The Complete Works of Aristotle*, Volume 1, Jonathan Barnes, ed (Princeton University Press, 1984)

[5] The formal language is many-sorted, and other styles of variable will be introduced presently

[6] Understood throughout to be governed by the principles of classical extensional mereology such as those of a first-order version (without sets) of Henry Leonard and Nelson Goodman, "The Calculus of Individuals and Its Uses," *Journal of Symbolic Logic*, v (1940) 45–55, or Peter Simons's system CI, *Parts A Study in Ontology* (New York Oxford, 1987), p 48

$$(2) \qquad \varphi(\pi) \wedge \varphi(\rho) \supset \varphi(\pi \cup \rho)$$

More generally, a predicate, φ, is cumulative iff

$$(3) \qquad \exists \pi\, \psi(\pi) \wedge \forall \pi(\psi(\pi) \supset \varphi(\pi)) \supset \varphi(\Sigma \pi\, \psi(\pi)),$$

where $\Sigma\pi\, \psi(\pi)$ is the general sum of all those things π satisfying ψ The existential clause, $\exists\pi\, \psi(\pi)$, stating that there is something satisfying the condition ψ, is necessary to ensure the existence of the sum (in accordance with the general mereological axiom for the existence of sums) Given this, if everything satisfying ψ is φ, then so is the sum of whatever satisfies ψ

Quine denies the general distributive condition, but suggests a restriction “In general, a mass term in predicative position may be viewed as a general term which is true of each portion of the stuff in question, excluding only the parts too small to count Thus, ‘water’ and ‘sugar’ are true of each part of the world’s water or sugar, down to single molecules but not to atoms” (*ibid* , p 98) As a restriction on the distributive condition, however, excluding *only* the parts too small to count will not work There are mereological sums of “parts too small to count,” such as the hydrogen and the oxygen in a quantity of water, which are not water but may well be larger than some things which are water Nevertheless, these “large” parts will be overlooked for the time being (they are taken up in the next section), and then it might be said that such counterinstances are not recognized from the macroscopic perspective [8] Quine’s proposal suggests that the cumulative and distributive conditions are independent, allowing the latter to be denied and perhaps replaced by some restricted version, while retaining the former It might be taken to indicate how a viewpoint integrating macroscopic and microscopic perspectives could be formed by modifying the former But the issue is not so straightforward

Water is an ordinary chemical substance with a disposition to react with other substances and a susceptibility to decomposition by physical methods such as heating and electrolysis It is also formed in many reactions The possibility of chemical transformation means that the property of being water is something possessed by a quantity of matter at one time or another Because of this, the predicate 'is water' is a two-place relational predicate, and we should write 'π is water at time t' when we need to be fully explicit From the macroscopic perspective, what is of relevance is what is true of intervals of time, our observational capacities and instrumental resources telling us what is true of longer or shorter intervals of time Treated as intervals, times stand, like quantities of matter, in mereological relations of parthood, overlapping, and separation, and are subject to the binary mereological operations of sum and product[9] So mass predicates can include relational predicates, and some of the mass predicates ordinarily treated as monadic predicates turn out, in fact, to be relational Further, a plurality of substances requires a distinction between quantities of matter which are, and those which are not, of the same substance Accordingly, 'is the same kind of substance' is another example of a relational mass predicate Any of the white solid covering much of Greenland is the same substance as the liquid, or rather what constitutes much of the liquid, in the Atlantic But this holds for a quantity of the white solid only while it remains water Most of what was water in the tank yesterday before being decomposed by electrolysis was the same substance then as the matter now in the tank is now Two quantities of the same kind are not in general indefinitely of the same kind, and the relation to time must be incorporated into a complete representation of the same kind relation, which is a four-place relation 'π is the same kind of substance at time t as ρ is at t''

---

[7] Quine, *Word and Object* (Cambridge MIT, 1960) p 91, where he mentions Goodman’s earlier use of the term ‘collective’ rather than ‘cumulative’

[8] Simons (p 155) refers to this perspective as logical analysis when he says “although we know that as a matter of fact there are smallest possible portions of gold, namely gold atoms, this is not something which could be known a priori and should not therefore have any part in the *logical* analysis of such predications ”

How should the distributive and cumulative conditions be modified to apply to relational predicates? The distributive condition is straightforward, it generalizes to

$$\varphi(\pi, t) \wedge \rho \subseteq \pi \wedge t' \subseteq t \supset \varphi(\rho, t') \tag{4}$$

in the case of a dyadic predicate $\varphi(\pi, t)$, and predicates of higher arity follow the same pattern[10] But the cumulative condition raises a problem Ignoring for simplicity the time-dependency of the 'same kind' predicate, a straightforward generalization of the cumulative condition (2) applied to this predicate would read

---

[9] Needham, "Temporal Intervals and Temporal Order," *Logique et Analyse*, xciii (1981) 49-64

[10] A referee asks why the mereological relation is not relativized to a time Given that the temporal qualification is put on the substance predicate, relativization of the mereological relation is not necessary what undergoes change of substance or phase properties does not thereby become a different quantity Unlike an individual such as a living organism, which undergoes change in its material constitution, the quantities of material comprising the organism's material constitution do not themselves change their parts

> (*) If π is the same substance as ρ and π' is the same substance as ρ', then $\pi \cup \pi'$ is the same substance as $\rho \cup \rho'$

But the antecedent would be true if π and ρ were each water and π' and ρ' each carbon dioxide, whereas the consequent is false Here all π is the same substance as all ρ, and all π' is the same substance as all ρ', but it is clearly not the case that all $\pi \cup \pi'$ is the same substance as all $\rho \cup \rho'$

Putting it like this may seem to make an appeal to the distributive condition that is questionable at this stage of the argument in view of Quine's objection But his objection only concerns micro-constituents below the molecular level These are not at issue in this counterexample because π is a quantity of macroscopic proportions and obviously a part of $\pi \cup \pi'$, and ρ' is similarly a macro-quantity and a part of $\rho \cup \rho'$, but water is not the same substance as carbon dioxide and so π is not the same substance as ρ' Consequently, it is not the case that all $\pi \cup \pi'$ is the same substance as all $\rho \cup \rho'$, and Quine's restriction would not save the suggested generalization of the cumulative condition In any case, were the consequent true, then reapplying (*) to it together with another same substance claim would entail that larger and more disparate sums were of the same substance, and eventually all distinctions of substance would be eliminated

This implies that the same substance predicate does not relate mixtures of several substances, but has the sense of 'same single substance' A more general 'same stuff' predicate would include the relation of being the same mixture, and perhaps (if some philosophers' usage is to be followed) be restricted to certain phases Such a general 'same stuff' predicate may not be a mass predicate in any sense distinguished below

Making the time-dependency of substance predicates explicit yields a similar counterexample, although in this case a condition ensuring the existence of the sum of two intervals must be added Thus,

$$(\dagger)\ Water(\pi, t) \wedge Water(\rho, t') \wedge Con(t, t') \supset Water(\pi \cup \rho, t \cup t'),$$

is not true in general The clause *Con*(t, t'), stating that t and t' are connected (abut or overlap), is necessary to ensure the existence of their sum Suppose that π is a quantity of water at t, and subjected to electrolysis during t', later than but abutting t, when it gradually dissociates into hydrogen and oxygen Then the sum, $\pi \cup \rho$, would not be water at $t \cup t'$, even if ρ is water during t' If the consequent of (†) were true, then reapplying (†) to it together with a further claim of a new quantity being water at a time connected with $t \cup t'$, and so on, would entail that ever larger quantities of shifting composition are water for ever larger intervals. Further counterexamples are to be had with 'is as warm as', 'is warmer than', 'has the same pressure as', and the time-dependent phase properties of being solid, liquid, and gas.

A satisfactory formulation of the cumulative condition must reduce to, and so be logically equivalent with, the general formulation (3) in the special case of monadic predicates. But the logically equivalent version for monadic predicates must be generalizable in an acceptable way. A solution has been put forward by Peter Roeper[11] to the effect that (3) is equivalent with

$$\forall \pi' \subseteq \pi \exists \pi'' \subseteq \pi' \varphi(\pi'') \supset \varphi(\pi) \tag{5}$$

(5) can be generalized into a suitable cumulative condition for relational predicates which is not subject to the counterexamples which tell against the likes of (*). For a dyadic substance predicate like 'water', for example, we have

$$\forall \pi' \subseteq \pi \forall t' \subseteq t \exists \pi'' \subseteq \pi' \exists t'' \subseteq t' \; Water(\pi'', t'') \supset Water(\pi, t) \tag{6}$$

That is, π is water throughout t if some of any part of π during some part of any subinterval of t is water.

But the proof of the equivalence, and more specifically, of the implication of (3) by (5), relies on the distributive condition (1). To show this, the proof is set out in an appendix. In the absence of some alternative deduction of (3) from (5) which does not require the unrestricted distributive condition, it would seem that the motivation for a cumulative condition like (6) for relational predicates would be hard to justify without the distributive condition. This falls short of saying that (5) implies (3). The important point is that the quantifiers in the antecedent of (6) range over all the parts of water without restriction, and the feature which Quine found objectionable in the distributive condition has come to roost in the cumulative condition. The domain of a theory expressing substance kinds by mass predicates is confined to macroscopic bodies, and there are no resources for describing microentities.

III

The distributive condition is still in trouble without appealing to parts smaller than any which are, for example, water. As suggested in the last section, the distributive condition is also subject to counterinstances in the form of parts of tangible quantities of water much more massive than the smallest quantities of water The oxygen composing such a tangible quantity of water is a case in point, as is the hydrogen It seems that confining the domain of quantification to the macrorealm as required by the cumulative condition is not enough to save the distributive condition Whether this constitutes a genuine counterexample depends, however, on the conception of mixture which water, as a combination of hydrogen and oxygen, is held to exemplify, and the corresponding understanding of parthood Two views of mixture are described here, each deriving from ancient sources On the Aristotelian view, water is construed so as not to offer a counterinstance to the distributive condition, whereas the Stoic view construes water along the lines just indicated The views differ on how the mereological relations between quantities stand to mereological relations between spatial regions This leads to a distinction between the general notion of parthood between quantities and the specific notion of spatial parthood on the Stoic view, and a corresponding distinction between distributive and cumulative conditions which leads in turn to a distinction amongst mass predicates between *intensive* and *homogeneous* predicates

---

[11] Roeper, "Semantics for Mass Terms with Quantifiers," *Noûs*, XVII (1983) 251–65, see p. 259.

The Aristotelian theory of mixture has it that the elements which go to make up a compound are no longer actually present in the compound This is expressed by Aristotle's distributive condition, which allows no part of a given compound which is not of that compound kind, just as there is no part of what he took to be the element water which is not water The elements are, however, potentially present in the compound The feature of not being actually present serves to contrast the Aristotelian theory with a view here called the Stoic theory of mixture On this conception, a substance can be separated from a mixture only if it is actually present in the mixture, although the Aristotelian view of the spatial uniformity of compounds, that every spatial part of a compound is of that same compound kind, is upheld Conflict is avoided by rejecting the Aristotelian view that two quantities of matter cannot occupy the same region of space at the same time, and holding that spatial uniformity is consistent with the elements composing a compound occupying the same region as the compound and as one another The Stoic view apparently circumvents the modal notion to which the Aristotelian view appeals But this will be reconsidered after the modal commitments of the two views are taken up in the next section

Applied to the post-Lavoisian conception of water as composed of hydrogen and oxygen, the Aristotelian view would have it that neither hydrogen nor oxygen are actually present in any quantity of water, although these elements are potentially recoverable from water The Stoic view has it that parts of water are hydrogen and others are oxygen, although these are not in general spatial parts It seems that the distributive condition holds on the Aristotelian view but not on the Stoic, the mereological relation of parthood being held to entail spatial parthood in the one case and not in the other But a notion of spatial homogeneity is upheld on the Stoic view, which can be distinguished as follows

Spatial parthood is a matter of parthood relations obtaining between the spatial regions occupied by quantities of matter Since occupying a region is a time-dependent notion—a body occupies one region at one time, but might move to another at another time—so is the corresponding notion of spatial parthood Writing $Occ(\pi, p, t)$ for 'π occupies region p for t', where $p, q, r,$ are variables ranging over spatial regions, this is understood to mean 'π occupies exactly the region p during t', so that it can also be written in functional form $Occ(\pi, t) = p$ Since cooccupancy is a possibility on the Stoic view, $Occ(\pi, p, t)$ does not necessarily imply that π exhausts the occupants of p on that view A relation which does have this implication can be defined by

$$\text{Def } MaxOcc(\pi, p, t) \equiv Occ(\pi, p, t) \wedge \pi = \Sigma\rho\ Occ(\rho, p, t)$$

On the Aristotelian view, $Occ(\pi, p, t)$ implies $\pi = \Sigma\rho\ Occ(\rho, p, t)$ and no distinction is made Since the region at issue is still exactly that maximally occupied, $MaxOcc(\pi, p, t)$ can still be written in functional form as $MaxOcc(\pi, t) = p$[12]

Now the Aristotelian maintains that cooccupancy is impossible, so parts are always spatial parts and the distributive condition holds on this view as stated[13] Not so for the proponent of the Stoic view, who accepts uniformity of substance in the sense

$$(7) \qquad \varphi(\pi, t) \wedge t' \subseteq t \wedge MaxOcc(\rho, t') \subseteq MaxOcc(\pi, t') \supset \varphi(\rho, t'),$$

where $\varphi(\pi, t)$ is a substance predicate, but not (4) To see whether a simple comparison with (4) is possible, note first that if ρ is the total

---

[12] It is natural to think of the region occupied by a quantity of matter at a time as the sum of the regions occupied by its parts during subintervals of that time Laying this down as an axiom,

$$Occ(\pi, p, t) \supset p = \Sigma q \exists\rho\exists t'(\rho \subseteq \pi \wedge t' \subseteq t \wedge Occ(\rho, q, t')),$$

the exactness of place occupied follows from the uniqueness of sums *MaxOcc* is then more appropriately defined by

$$MaxOcc(\pi, p, t) \equiv \pi = \Sigma\rho\ \exists q \subseteq p \exists t' \subseteq t\ Occ(\rho, q, t')$$

[13] If spatial occupancy introduces a time factor, as the above development of the Stoic view suggests, then the claim just made in the text would seem to presuppose that the Aristotelian view subscribes to the equivalence

$$\rho \subset \pi \equiv \forall t \exists p \exists q\ (Occ(\rho, p, t) \wedge Occ(\pi, q, t) \wedge p \subset q)$$

material content of a region at one time and this region is a proper part of the region exhaustively occupied by π at this *same time*, then ρ is a proper part of π.

$$MaxOcc(\rho, t) \subset MaxOcc(\pi, t) \supset \rho \subset \pi$$

The converse is not in general true on the Stoic view. Moreover, quantities maximally occupying identical regions are identical. Thus,

$$MaxOcc(\rho, t) \subseteq MaxOcc(\pi, t) \supset \rho \subseteq \pi$$

For the purposes of comparing (7) with (4), what must be considered is the more general claim

$$t' \subseteq t \wedge MaxOcc(\rho, t') \subseteq MaxOcc(\pi, t) \supset \rho \subseteq \pi, \tag{8}$$

where the times are not necessarily the same, but also stand in the part relation. This generalization does not change matters as far as I can see, however, and (8) holds too. I do not have a general proof to offer, but it may be instructive to see why an attempted counterexample does not succeed. Suppose, then, that ρ is an aqueous solution of which π is the water content exactly filling a sturdy container with a movable lid during some initial subinterval t' of t, and the lid is raised like a piston later on in t so that some of the water in ρ evaporates, entering the vapor phase with the result that ρ occupies a larger region encompassing that which it originally occupied during t'. The water, π, will occupy at t the sum of the regions occupied at subintervals of t, and this will include as a proper part that occupied by the solution during the subinterval t' of t. But although ρ and what is the solution at times after the raising of the piston are not parts of π, π does not maximally occupy the region in question.

From (8) it follows that the antecedent of (7) is in general stronger than that of (4), making (7) the weaker principle on the Stoic view.

The ancients seem not to have considered the cumulative condition, but an analogous restriction to spatial parts can be introduced there, so that instead of (6) we have

$$\begin{aligned} &\forall \pi' \forall t' \exists \pi'' \exists t'' \, (MaxOcc(\pi'', t'') \subseteq MaxOcc(\pi', t') \subseteq MaxOcc(\pi, t) \\ &\quad \wedge t'' \subseteq t' \subseteq t \supset \varphi(\pi'', t'')) \supset \varphi(\pi, t) \end{aligned} \tag{9}$$

The universal quantifier cannot be placed after the existential quantifiers because quantities can move and change their position. Dropping the universal quantifier would give a 3-place, time-dependent relation. But presumably, a part at any time is a part at every time on the Aristotelian view. At any rate, there is no time at which a sometime part is not a part.

The antecedent of the main conditional connective is itself a conditional sentence, whose antecedent is, by (8), stronger than the corresponding antecedent of (6) Consequently, this antecedent conditional is weaker, making the main conditional *stronger* than that of (6) Let us say that predicates satisfying both the distributive and cumulative conditions in the original form (4) and (6) (for whatever sort of variables) are *intensive*, and those satisfying both the explicitly spatial interpretations of these conditions, (7) and (9), are *homogeneous* On the Aristotelian view this is a distinction without a difference, but not on the Stoic view The upshot is that, on the Stoic view, the classifications of predicates as intensive and homogeneous cut across one another, since neither of these features implies the other Accordingly, mass predicates can be divided into three categories, those which are both homogeneous and intensive, those which are homogeneous but not intensive, and those which are intensive but not homogeneous Nonintensity means either not (4) or not (6), but since (9) implies (6), not (4) is the only possibility in conjunction with homogeneity Similarly, since (4) implies (7), being intensive but not homogeneous amounts to satisfying (4), (6), (7) and not (9)

To which category does the relational 'same substance' predicate belong? As we have seen, on the Stoic view compound substance predicates like 'water' are homogeneous but not intensive The corresponding 'same substance' predicate, more specifically the 'same compound' predicate, is accordingly homogeneous but not intensive But elemental substances are both homogeneous and intensive, and the corresponding 'same substance' predicate, more specifically the 'same element' predicate, is both homogeneous and intensive Further, in classical (nineteenth century) chemistry where the elements are thought of as nontransmutable, an elemental substance predicate, such as 'is oxygen', can be treated as a monadic predicate There is no reason to treat it as a dyadic time-dependent predicate if, unlike water, oxygen never ceases to be oxygen, and what is not oxygen never becomes oxygen The corresponding 'same element' predicate is therefore also time-independent, and so a dyadic predicate of quantities[14] This difference is dropped in twentieth-century chemistry, which recognizes that the elements were once created by intra-stellar

---

[14] The spatial distributive condition for the "same element" predicate, *SE*, would then have a spatial parthood relation instead of the parthood between times of (7), thus

$$SE(\pi, \rho) \wedge MaxOcc(\pi', t) \subseteq MaxOcc(\pi, t) \wedge MaxOcc(\rho', t) \subseteq MaxOcc(\rho, t) \supset SE(\pi', \rho')$$

and on this view of the elements, it would be reasonable to maintain that once a spatial part, always a spatial part, that is, $MaxOcc(\pi', t) \subseteq MaxOcc(\pi, t) \supset \forall t\, MaxOcc(\pi', t) \subseteq MaxOcc(\pi, t)$ as a general principle Thermal expansion in the gas phase may well lead fusion processes and are in some cases destroyed by radioactive disintegration Considered as a homogeneous predicate, however, the disjunctive character of the sameness of substance concept is exclusive Being either the same element or the same compound is not like being liquid or gas A quantity is liquid or gas if it is all liquid or gas, that is, any spatial part is either all liquid, all gas, or partly liquid and partly gas Two quantities of matter are the same substance if they are each all the same element or each all the same compound, there is no question of them being partly the same element and partly the same compound

The three categories of mass predicate are not distinguished on the Aristotelian view, and the conception of elements, and how they are contrasted with compounds, is altogether different Aristotle's own scheme distinguished substances by their degrees of the comparative qualities of being warmer or colder than something, and being moister or dryer than something He thought that there were maximal degrees of these comparative qualities, which he called "contrary extremes" (*DG* II 8, 335^{a}8)—what is as warm as can be is Hot, what is as cold as can be is Cold, and similarly for Moist and Dry The general notion of an element is that of a substance possessing compatible contrary extremes Particular elements were characterized by the particular pair of compatible extremes which they display in isolation (water is Cold and Moist, earth is Cold and Dry, and so on) but lose when brought together in a mixing process which therefore destroys the elements and creates a compound in their place Loss of these features means, on the Aristotelian view, that there are no elements in compounds because no part of what is a compound bears the defining characteristics of any of the elements, which therefore only occur in isolation

The macroscopic notion of a phase property, such as 'is liquid', is a time-dependent notion A given quantity of water may be liquid at one time but gas or solid under conditions obtaining at other times A quantity of water comprising ice floating on the liquid is a two-phase mixture containing just one chemical substance Aristotle seems to have thought that the notions of phase and substance go hand in hand Transformation of liquid water to solid or gas would be a transformation of one kind of substance to another A two-phase mixture would necessarily be a mixture of two substances Preservation of

---

to some part of a quantity of an element occupying a more inclusive volume at one time than that quantity at another time But the principle in question speaks of the same time Similarly for the spatial cumulative condition phase is, on the Aristotelian view, a necessary condition of preservation of substance kind (and Putnam seems to follow Aristotle when he claims that determining whether something is the same kind as a quantity of water is a matter of being the same liquid) But conditions under which a quantity of a single substance can comprise parts of different phase are described by Gibbs's phase rule, an important theorem of thermodynamics which can be used (amongst other things) to identify the characteristic behavior of a quantity comprising just one substance What are called *the* phases in a mixture of ice and liquid water are the maximal quantities of liquid and of solid—the quantity which includes all and only the liquid and the quantity which includes all and only the solid Now all the spatial parts of what is liquid for a given time *t* are liquid for all subintervals of *t*, and any quantity is liquid for *t* if any spatial part during any subinterval of *t* has a spatial part for a subinterval of that subinterval which is liquid Phase properties are thus homogeneous, but they are also intensive If there is any hydrogen forming part of the water that runs out of the tap, then it is liquid, as is any carbon dioxide when dissolved in chloroform at normal temperature and pressure The properties of warmth (temperature) and pressure, and the corresponding relational properties of being as warm as, warmer than, having the same pressure as and having higher pressure than, are also homogeneous and intensive

The Stoic view therefore puts elemental properties and phase properties in the same category There is no implication that they cannot be distinguished, just that the distinction is not made by the distinction of categories of mass predicates A given quantity of matter might well be all of a single elemental kind, and yet not all in a single phase, in the extreme case, a mixture of three distinct phases is possible, in accordance with the phase rule, at the triple point And air, for example, is a solution containing several elements in the same phase Similarly, on this Stoic conception, predicates describing compounds and solutions belong to the same category of mass predicates, being homogeneous but not intensive This distinction was apparently not known to the ancients, who had no conception of the principle of constant proportions which served as the simple criterion differentiating uniform mixtures into these two kinds from the beginning of the nineteenth century[15] Modern proponents of what is here called the Stoic view would consider the components of a solution to be actually present in the mixture as distinct parts, just as the elemental constituents are distinct parts of a compound So solutions are nonintensive although homogeneous But with the recognition of Berthollide compounds of the transition elements, whose composition, like that of a saturated solution, varies with circumstances, the distinction between compounds and solutions is not as clear-cut as it was once thought to be Since compounds and solutions belong to the same category, the classification of mass predicates would not be affected should the distinction between them become blurred

---

[15] But see the discussion of "same ratio" following (11) below

There is much in the Stoic view that accords with modern ways of talking about substances in mixtures, be they heterogeneous, solutions, or compounds Does the cooccupancy claim detract from this? In thermodynamics, the energy of a multi-component system subject to work in the form of change in volume is specified as a function $U(S, V, N_1, \quad , N_r)$ of the entropy, S, the relative amount of each component substance, N_i, and the volume, V Although the amounts of the several kinds of component substances are specified, only one term for the volume, that of the whole system, appears This looks very much like cooccupancy It was certainly the view of Pierre Duhem, who studied the matter in detail [16] And modern textbook writers find it natural to speak of "subsystems which occupy the same region of space "[17]

Philosophers have balked at the idea, claiming, for example, that "impenetrability implicitly defined by the principle that no two things can be in the same place at the same time is entailed by the occupancy of space "[18] But arguments for this stance are thin on the ground Wiggins claims that distinct objects of the same general kind in the same place cannot be distinguished because "no volume or area of space can be qualified simultaneously by distinct predicates in any range (color, shape, texture, and so forth) "[19] But a gaseous solution such as the air exhibits the different partial pressures of each component substance Thermodynamics ascribes to each component substance in a mixture its own chemical potential, and they may well have different concentrations (mole fractions) and degrees of ionization

From the macroscopic perspective, quantities of matter occupy regions of space for intervals of time Macroscopic ontology is treated here as in "[t]hermodynamics," which is standardly interpreted to be " independent of the fine structure of matter "[20] The macroscopic perspective is thus neutral toward the details of any view of the nature and behavior of discrete microstructures ascribed to macroscopic quantities on a microscopic perspective According to the microscopic perspective, these discrete particles are thought of as moving very fast and, in so far as they can be located, sweeping out a region of space in any given interval of time, as the result of vibrational and rotational motion, and translational motion in fluids Even on the unrealistic view of microparticles as impenetrable bodies with definite bounding surfaces, they will sweep out regions which will overlap even during the shortest intervals of macroscopic significance If cooccupancy is a problem, it is only for localized microparticles with well-defined boundaries at instants of time Modern chemistry, of course, entertains no such entities, and considers the character of chemical species over different time spans, to use Steven Weininger's[21] expression But the macroscopic perspective is neutral on the details of microscopic composition, and the stand on cooccupancy implied by the Stoic view as understood here is consistent even with this extraordinary view of the microworld

---

[16] Needham, "Duhem's Theory of Mixture in the Light of the Stoic Challenge to the Aristotelian Conception," *Studies in History and Philosophy of Science*, XXXIII (2002) 685–708, see pp 705–07

[17] Roger Bowley and Mariana Sánchez, *Introductory Statistical Mechanics* (New York Oxford, 1999), p 189

[18] Anthony Quinton, "Matter and Space," *Mind*, LXXIII (1964) 332–52, see pp 342–43

[19] David Wiggins, "On Being in the Same Place at the Same Time," *Philosophical Review*, LXXVII (1968) 90–95, see p 94

IV

The Aristotelian notion of the potential presence of elements in a compound explicitly introduces a modal dimension into the understanding of substance concepts Applied to a quantity, π, of water, understood as formed from hydrogen and oxygen, it claims something along the lines of 'π is partly possibly hydrogen and partly possibly oxygen', which might be rendered

$$(10) \qquad W(\pi, t) \supset \exists\rho\exists t'(\rho \subset \pi \wedge \Diamond(H(\rho, t') \wedge O(\pi - \rho, t'))),$$

where $W(\pi, t)$ abbreviates 'π is water at t',$H(\rho, t)$ abbreviates 'ρ is hydrogen at t',$O(\rho, t)$ abbreviates 'ρ is oxygen at t' and '$-$' is the operation of mereological difference Whether a conjunct saying t is separate from (has no part in common with) t' should be introduced within the scope of the existential quantifiers is left open (10) is still not satisfactory as it stands, however But rather than modifying it, perhaps the starting point should be 'π is possibly partly hydrogen and partly oxygen', and the modal operator in (10) moved to the immediate left of the quantifiers This point is addressed below The notion of potentiality may well express something stronger than mere possibility, but that issue is not pursued here The immediately pressing question following the discussion above is what, if water really is distributive (as maintained on the Aristotelian view) and all the parts of a quantity of water are alike, could distinguish the part which might become hydrogen from that which might become oxygen?

---

[20] Kenneth Denbigh, *The Principles of Chemical Equilibrium* (New York Cambridge, 1981, 4th ed), p 4

[21] Weininger, "Butlerov's Vision The Timeless, the Transient, and the Representation of Chemical Structure," in Bhushan and Rosenfeld, eds , *Of Minds and Molecules*, pp 143–61, see p 147

This question does not really make sense unless parts are preserved across modal contexts Other conceptions raise other questions, but that is the mereological conception common to the two conceptions of mixture compared here It accords with the earlier discussion if it is granted that quantities never gain or lose parts because they cannot To make this fully explicit, it is laid down that a quantity necessarily preserves its structure of parts

(R⊆) (i) $\pi \subseteq \rho \supset \Box (\pi \subseteq \rho)$ and (ii) $\pi \nsubseteq \rho \supset \Box (\pi \nsubseteq \rho)$

Since the identity of quantities is determined by their parts, it will transpire that under reasonable modal assumptions these conditions imply the rigidity conditions (R)

Surely, in answer to the question, there are no appropriate distinguishing features and there should not be any unique division of a quantity of water into two parts, one of which is possibly hydrogen while the other is possibly oxygen Aristotle seems to have had some such thought in mind when he argued against the idea of a compound held by those who "conceive it as *composition*—just as a wall comes-to-be out of bricks and stones, and this mixture will be composed of the elements, these being preserved in it unaltered but with their small particles juxtaposed each to each " He continues

> That will be the manner, presumably, in which flesh and every other compound results from the elements Consequently, it follows that Fire and Water do not come-to-be out of any and every part of flesh For instance, although a sphere might come-to-be out of *this* part of a lump of wax and a pyramid out of *some other* part, it was nevertheless possible for either figure to have come-to-be out of either part indifferently *that* is the manner of coming-to-be when both come-to-be out of any and every part of flesh (*DG* II 7, 334^{a}29ff)

Just as a sphere and a pyramid can be obtained from *any* way of dividing a lump of wax into two parts of given relative size, so it must be possible to derive the elements from "any and every part of flesh " The Aristotelian conception of the potential presence of elements in compounds is one in which a given division into elements is possible, but does not exhaust the possibilities

There is a question of how to formulate this condition Immediately preceding the passage from *DG* I 10 quoted at the beginning of section II where Aristotle says that a compound must be uniform, he also requires that "the part exhibit the same ratio between its constituents as the whole" (328^{a}8–9) Any other division preserving the ratio of the elements is equally possible

$$(11)\quad W(\pi, t) \supset \exists\rho\exists t'(\rho \subset \pi \wedge \Diamond(H(\rho, t') \wedge O(\pi - \rho, t')) \wedge \forall\sigma(\sigma \subset \pi \wedge \mathit{SameRatio}(\rho, \pi - \rho, \sigma, \pi - \sigma) \supset \Diamond(H(\sigma, t') \wedge O(\pi - \sigma, t'))))$$

'$\mathit{SameRatio}(\pi, \rho, \sigma, \tau)$' here abbreviates the relation of the ratio of the amount of matter π to that of ρ being the same as that of σ to τ Quantitative features are controversial in the interpretation of Aristotle But without some restriction on the universal quantifier akin to that expressed by the 'same ratio' clause here, the Aristotelian view of potential presence would be distinctly less attractive[22]

The Aristotelian view will not strike the modern reader as particularly plausible The Stoic view, which keeps track of the part of water exhibiting the characteristic features of hydrogen in isolation (density, boiling point, and so on) before and after its involvement with water, and similarly for oxygen, seems more reasonable This view is not without its modal commitments, however The Stoic view is not opposed to (10), but instead of (11) would assert the uniqueness of the parts at issue, even the necessary uniqueness

$$(12)\quad W(\pi, t) \supset \exists\rho\exists t'(\rho \subset \pi \wedge \Diamond(H(\rho, t') \wedge O(\pi - \rho, t')) \wedge \Box\forall\sigma (\sigma \subset \pi \wedge \Diamond(H(\sigma, t') \wedge O(\pi - \sigma, t')) \supset \rho = \sigma)),$$

where $\rho = \sigma$ in the last conjunct is taken to imply $\pi - \rho = \pi - \sigma$ The Stoic view in fact goes further than (10), and asserts the corresponding claim with the possibility operator dropped, similarly for (12) Definite ratios of elements in compounds is a matter of comparison between quantities of the same compound, and both views can agree that any quantities of water have the same ratio of elements

$$(13)\quad W(\pi, t) \supset \exists\rho(\rho \subset \pi \wedge \exists t'\Diamond(H(\rho, t') \wedge O(\pi - \rho, t')) \wedge \Box\forall\sigma\forall t'\forall\tau\forall t''(W(\sigma, t') \wedge \tau \subset \sigma \wedge \Diamond(H(\tau, t'') \wedge O(\sigma - \tau, t'')) \supset \mathit{SameRatio}(\rho, \pi - \rho, \tau, \sigma - \tau)))$$

---

[22] I give a speculative interpretation of this relation along Aristotelian lines in "Aristotle's Theory of Chemical Reaction and Chemical Substances," in Davis Baird, Eric Scerri and Lee McIntyre, eds , *Philosophy of Chemistry Synthesis of a New Discipline* (Boston Springer, 2006), pp 43–67

The rigidity of the singular terms is essential for the expression of these views. Clearly, π must refer to the same thing in all its occurrences, both inside and outside the modal operators. The necessary invariance of composition as expressed by (13), and the uniformity expressed by (11), also presuppose that the measure of the amount of matter comprising a quantity, and the relation of *SameRatio* on which it is based, cannot vary. When the relation holds, it is necessarily preserved, in fact, it satisfies the following analogues of the rigidity conditions.

(Rratio) $\mathit{SameRatio}(\pi, \rho, \sigma, \tau) \supset \Box\mathit{SameRatio}(\pi, \rho, \sigma, \tau)$ and
$\sim\mathit{SameRatio}(\pi, \rho, \sigma, \tau) \supset \Box \sim\mathit{SameRatio}(\pi, \rho, \sigma, \tau)$

If such ratios cannot vary, they will not be able to differ from one time to another, and so the *SameRatio* predicate is not time dependent with place-holders for time variables. This stands in contrast to the treatment of the 'same substance' predicate.

Both Aristotelian and Stoic views hold that compound properties such as being water are not necessary features of the things that bear them. (It might be thought not possible that what is water at a given time is not water at this same time, as a feature of time-dependent predicates in general rather than any specific feature of substances. But subjunctive conditionals of the kind 'If this water had just been subjected to electrolysis, it would now be decomposed into its elements' argue against this general thesis.) In general, writing $S(\pi, t_1, \rho, t_2)$ for 'π is the same substance at t_1 as ρ is at t_2', being the same substance is not necessarily preserved and does *not* satisfy the analogue of the first of the rigidity conditions (R) described in section I,

(RSi) $S(\pi, t_1, \rho, t_2) \supset \Box S(\pi, t_3, \rho, t_4)$,

nor any special case involving identification of the time variables. Nor does it satisfy the other analogue,

(RSii) $\sim S(\pi, t_1, \rho, t_2) \supset \Box \sim S(\pi, t_3, \rho, t_4)$

If π is water and ρ is a mixture of two volumes of hydrogen with one of oxygen, they are not the same substance (π is liquid at normal temperature and pressure, ρ is not, and when both are gases, π has two thirds the volume at the same temperature and pressure) though they could be. Or if π and ρ are isomers with the same compositional formula (deemed different substances because, say, of a difference in some sharp physical property such as boiling point or rotatory power), then we have a counterexample. Now the nineteenth-century Stoic view, we saw, holds the elements to be incorruptible, and thus exceptions to the present rule. Unlike compound substances which may be transformed into their components or into other substances, they are preserved while composing compounds and not transmutable into one another or anything else But the twentieth-century Stoic view recognizes that elements decompose by radioactive decay No chemical substances constitute essential properties of any matter, and the only essential features of quantities seem to be general ones such as always occupying some place— $\Box\, \forall t \exists p\ Occ(\pi, p, t)$

Although nothing is necessarily water, nothing could be water without being composed solely of hydrogen and oxygen And conversely, nothing could be H_2O without being water The composition of water, as distinct from hydrogen peroxide, is given by the compositional formula, H_2O This can also be written OH_2 (and called oxygen hydride as well as hydrogen oxide) since the compositional formula says nothing about structure, but merely states that hydrogen and oxygen are combined in the proportion of two chemical equivalents to one Were there isomers of water, with this same compositional formula, it would not be generally true that all H_2O is water, just as some, though not all, C_2H_6O is alcohol But since water has no isomers the equivalence between being water and being H_2O holds as stated

There seems to be broad agreement that this equivalence holds of necessity The claim is clearly compatible with the claim in the preceding paragraph that the 'same substance' predicate does not satisfy the analogues of the rigidity conditions, which precludes that anything is necessarily water, or necessarily H_2O, but not that necessarily whatever is water is H_2O and conversely As distinct from the 'same substance' predicate, relating quantities and times, this latter condition just gives a characterization of the 'is water' predicate by providing another which can be substituted for it in modal contexts Many other predicates serve in the same way, a good example is that expressing the triple point of water (0 01 °C and 4 58 mm Hg)—that a quantity of the pure substance at equilibrium at this temperature and pressure is partly solid, partly liquid, and partly gas

Although solutions are usually described by names of their constituents, such as a solution of copper and nickel, there are a few expressions in common use for kinds of matter which are solutions Brine, for example, is a solution of salt and water, and brass is a solution of copper and zinc These instantiate a schema along the lines of necessarily, for specified conditions of temperature and pressure, material at equilibrium has composition falling within specific ranges iff it is a certain kind of solution Compounds called Berthollides are, in a sense, intermediate cases Their composition is specified by a given formula, albeit a nonstoichiometric one (in which the indices are nonintegral), but only for a fixed temperature and pressure. Ordinary usage contains many other terms which philosophers[23] may be pleased to call kinds of stuff, with tighter or looser criteria of application, but which do not apply to single substances. Advertising hype calls the recipe for *Coca Cola* a secret formula. But the specification of a solution as a kind of stuff iff its components fall within very narrow ranges of concentration, without any reference to the laws governing the chemist's use of the term 'formula', is just an arbitrary stipulation which can be, and I believe has been, changed. Following the same line of thought, fogs and smogs of the past have no doubt satisfied certain general conditions with varying chemical constitution, and there is no saying how the chemical constitution of future smogs will continue to vary. Failure of terms from this assorted ragbag to stand in necessary equivalences detailing a definite chemical formula does not show that terms appropriated to name matter satisfying systematic (thermodynamic or other) criteria for being a single substance fail to do so. There are, in fact, "chemical kinds" (discussed in physical chemistry textbooks) which are not characterized by component substance kinds. Colloids, for example, may take the form of smokes, foams, emulsions, gels, sols, and so on, containing small (but macroscopic) particles of a dispersed phase (which may be liquid or solid) in a dispersing phase (gas, liquid, or solid). Necessarily, whatever is smoke is a heterogeneous mixture of small solid particles dispersed in a gas.

Another question raised by (10) is the order of the existential quantifiers with respect to the possibility operator. Would there be any difference if the above discussion started with

(10a) $W(\pi, t) \supset \Diamond\exists\rho\exists t'(\rho \subset \pi \wedge H(\rho, t') \wedge O(\pi - \rho, t'))$

rather than (10)? Rigidity ensures that π maintains the same reference in its occurrences outside and within the scope of $\Diamond$. But can the existential quantifier introduce new possible but nonactual quantities?

The general case of the relation between the existential and possibility operators raises the question of the Barcan formula, $\exists\pi\Diamond\varphi \supset \Diamond\exists\pi\varphi$, according to which quantities which actually exist cannot fail to do so, and the converse Barcan formula, $\Diamond\exists\pi\varphi \supset \exists\pi\Diamond\varphi$, according to which there is no possibility that quantities which might exist actually do not. For quantification over times and regions, these seem to be very natural principles indeed. As for quantities, it is natural to recall Lavoisier's classical principle of the constancy of mass during chemical transformation. This has usually been understood to express an idea of the noncreatability and indestructibility of matter (that is, the conservation of matter) after the introduction of mass as the universal measure of the amount of matter in Newtonian mechanics.[24] In the present context, this is naturally understood in terms of a principle of the conservation of quantities jointly expressed by the Barcan formula and its converse, namely that it is not possible that quantities which actually exist may not, nor that quantities may exist which actually do not.

---

[23] For example, Jerry Fodor, Review of Christopher Hughes, *Kripke Names, Necessity and Identity*, *London Review of Books*, XXVI (October 21, 2004) 17–19.

Lavoisier's conservation thesis does not, it seems, entail that the elements endure (are necessarily preserved), though he and nineteenth century chemists thought they did.[25] It seems, however, that transmutation of elements in the modern sense conflicts with the conservation principle as classically understood: there is a certain seepage of old matter with radioactive disintegration and the concomitant loss of binding energy, and a trickle of new matter with nuclear fusion. The classical conception of the conservation of mass is not strictly true, then, holding only as an excellent approximation for virtually all chemical and biochemical processes under accessible conditions. But the principle of the conservation of quantities may be able to accommodate these effects of transmutation by understanding quantities to be bearers of the property of mass-energy rather than simply mass. If this is acceptable, then (10) and (10a) would be equivalent, quite apart from considerations of the rigidity of parts, in virtue of a quantifier shift law encompassing the Barcan formula and its converse applied to quantity variables as well as space and time variables. Similarly, necessarily whatever is water iff it is H_2O implies that being water iff being H_2O is an essential feature of matter.

Which principles of sentential modal logic are appropriate for this modality? Given the basic principles of normal modal logic (system K), the rigidity conditions (Ri) and (Rii) follow from ($R\subseteq$ i) and ($R\subseteq$ ii) and the mereological identity axiom, $\pi \subseteq \rho \wedge \rho \subseteq \pi \supset \pi = \rho$. The proofs, given in the appendix, rely on the law that $\Box$ distributes over $\wedge$ as an equivalence; $\Diamond$ does not, and $\Diamond\varphi \wedge \Diamond\psi \supset \Diamond(\varphi \wedge \psi)$ is invalid. This can be illustrated in Aristotelian terms.[26] Earth, for example, may be worked up into wood and into clay, but presumably it is not possible for what is earth at one time to be both wood and clay at some one other time Similarly, on the Stoic conception, with a modern example, a quantity π which is hydrogen at t might be possibly part of some water and possibly part of some methane (M), that is, such that

---

[24] Lavoisier's own formulation reads "in all operations of art and nature, nothing is created; an equal quantity of matter exists both before and after the experiment, ..."—*Elements of Chemistry*, by Robert Kerr, trans. (New York: Dover, 1965), p. 130.

[25] Lavoisier continues the passage in the last footnote: "the quality and quantity of the elements remain precisely the same, and nothing takes place beyond changes and modifications in the combination of these elements."

[26] Cf. Mary Louise Gill, *Aristotle on Substance: The Paradox of Unity* (Princeton: University Press, 1989) p. 150.

$$\Diamond\exists\rho W(\pi\cup\rho, t) \wedge \Diamond\exists\rho M(\pi\cup\rho, t)$$

But this does not imply

$$\Diamond(\exists\rho W(\pi\cup\rho, t) \wedge \exists\rho M(\pi\cup\rho, t))$$

What of the converse principle, which does hold in K? Gill (*ibid*) argues that it is not possible for earth to be (the material of) a statue because it is dry and crumbly, and does not retain the shape into which it is molded If it is first worked up into bronze, however, which does have the appropriate dispositions, then the production of a statue is possible Thus, some quantity, π, which is earth might be both (part of) the material of a statue and bronze, and so (in accordance with the principle now at issue), it might be the material of a statue This state of affairs is not in conflict with what Gill says since the conclusion, that π might be the material of a statue, carries no implication that π might be both the material of a statue and (yet remain) earth What, following Gill, is not possible is that some quantity be both the material of a statue at some time and earth at the same time Observing the distinction between subject and predicate in this way, it is difficult to see that K gets the combination of modality and sentential connectives wrong

Moving on to consider stronger principles, it seems difficult to deny the characteristic schema of system T, namely $\varphi \supset \Diamond\varphi$ Admittedly, actuality is often contrasted with potentiality, as though potentiality presupposes deprivation Some such distinction is readily accommodated by understanding the contrast to stand between what is actual and what is *merely* possible, with the latter defined along the lines of $\sim\varphi \wedge \Diamond\varphi$ But manifestations of, say, dissolving in water do not imply absence of the potential to do so As Gill puts it, potentialities are grounded in the nature of the subject, which endows them with the appropriate dispositions, and not in a state of deprivation (*ibid*, pp 179–80) Beyond this, the question is whether iterated modalities have any special significance or should be eliminated, along with nested modalities, in favor of the innermost modality as in S5 I am inclined towards the latter Accordingly, to say, for example, that water necessarily can boil at 100°C, $\Box(W(\pi, t) \supset \Diamond B(\pi, t))$, is equivalent to saying $\Box\sim W(\pi, t) \vee \Diamond B(\pi, t)$, or possibly being water implies possibly boiling at 100°C With these modal principles, (R$\subseteq$ ii) follows from (R$\subseteq$ i)

V

On the Aristotelian view, the elements are not actually present in a compound. They are characterized by features they exhibit only in isolation, and by the distributive condition, no part of the compound has these features. The Stoic view rejects this, and maintains that the elements are present in compounds. This underlies the reason for maintaining (12). Both views are agreed on the conservation of the parts involved. What distinguishes them is the substance properties they ascribe to the parts when the whole is a compound. Although Aristotle's actual definitions of the elements are antiquated, the general manner of characterizing elements by listing properties exhibited in isolation along with the kinds of compounds they can form does resemble what is found in a modern textbook description of the elements. This raises the question of how the elements are to be characterized on the Stoic view, and given, as hitherto suggested, that this view corresponds more closely to the modern conception of substance, how modern descriptions of the elements are to be understood. It seems doubtful that elemental properties can be defined, or characterized by a range of criteria, entirely in terms of nonmodal properties. Being oxygen, for example, is having the disposition to be a gas when not chemically combined and above –183°C at standard pressure, and to be more stable as a component of water when mixed with hydrogen at standard temperature and pressure. No quantity of oxygen could actually display both features at the same time. The very idea of keeping track of the parts of water exhibiting the characteristic features of hydrogen and oxygen in isolation before and after their combination in water suggests that what is elemental stuff exhibits the potential, both to display these characteristic features of the isolated state when in a state of combination, and to display the characteristic features of its compounds when in isolation. The Aristotelian view incorporates modality into its conception of mixture, and gives a nonmodal account of the features characterizing the elements. The Stoic view does not circumvent modal notions altogether, but incorporates them into the characterization of the elements.

The two views of mixture were introduced as alternative responses to the distributive condition on mass predicates, the Aristotelian view preserving this condition and the Stoic view rejecting it. Spatial homogeneity is retained on the latter view, however, by accepting the possibility of cooccupancy, which it was argued can be understood as a reasonable principle of macroscopic ontology. In general, mixtures comprise different substances or different phases (the "or" being inclusive). The traditional Aristotelian view that phase and substance coincide remains the vulgar conception, but this feature need not be incorporated into the Aristotelian view presented here It is this general conception of mixture which sanctions the necessary equivalence between being water and being H_2O, which I have contrasted with the nonnecessary relations of being the same substance and its negation

PAUL NEEDHAM

University of Stockholm

APPENDIX A

The proof that (5) implies (3) follows by the deduction theorem from the deduction of the consequent of (3), $\varphi(\Sigma\pi\ \psi(\pi))$, from its antecedents and (5), which can be set out as follows

(i) $\exists\pi\ \psi(\pi)$ [Assumption]
(ii) $\forall\pi\ (\psi(\pi) \supset \varphi(\pi))$ [Assumption]

Put $\pi = \Sigma\pi'\ \psi(\pi')$, which exists since by (i) there are ψ-ers Now by general mereology, anything which is a part of the sum of ψ-ers overlaps some ψ-er, that is,

(iii) $\forall\pi' \subseteq \pi\, \exists\sigma \subseteq \pi'\ (\psi(\sigma) \wedge \pi'\ \mathbf{o}\ \sigma)$

But by the definition of overlap, the conjunct $\pi'\ \mathbf{o}\ \sigma$ here implies that there is a part, π'', common to π' and σ, and by the distributive condition (1), the first conjunct then implies $\psi(\pi'')$ Accordingly,

(iv) $\forall\pi' \subseteq \pi\, \exists\pi'' \subseteq \pi'\ \psi(\pi'')$

Then, in view of (ii),

(v) $\forall\pi' \subseteq \pi\, \exists\pi'' \subseteq \pi'\ \varphi(\pi'')$
(vi) $\varphi(\pi)$ [from (v) by (5)]

Since $\pi = \Sigma\pi'\ \psi(\pi')$, this is $\varphi(\Sigma\pi'\ \psi(\pi'))$ Removing the assumptions (i) and (ii) yields (3) Q E D

Turning now to the proofs of rigidity of quantities, it is shown that, given the principles of modal system K, the rigidity conditions (Ri) and (Rii) applied to quantity variables follow from the rigidity of their parts, conditions ($R\subseteq$ i) and ($R\subseteq$ ii) Thus, the mereological identity axiom, $\pi \subseteq \rho \wedge \rho \subseteq \pi \supset \pi = \rho$, yields

(i) $\Box(\pi \subseteq \rho \wedge \rho \subseteq \pi) \supset \Box(\pi = \rho)$ [Necessitation, distribution of $\Box$ over $\supset$]
(ii) $\Box(\pi \subseteq \rho) \wedge \Box(\rho \subseteq \pi) \supset \Box(\pi = \rho)$ [distribution of $\Box$ over $\wedge$]
(iii) $\pi \subseteq \rho \wedge \rho \subseteq \pi \supset \Box(\pi \subseteq \rho) \wedge \Box(\rho \subseteq \pi)$ [($R \subseteq$ i), sentential logic]
(iv) $\pi = \rho \supset \pi \subseteq \rho \wedge \rho \subseteq \pi$ [Leibniz's law, $\pi \subseteq \pi$, sentential logic]
(v) $\pi = \rho \supset \Box(\pi = \rho)$ [(iv), (iii), (ii)]

Similarly, (Rii) follows from (R$\subseteq$ ii), again the mereological identity axiom yields

(vi) $\pi \neq \rho \supset \pi \nsubseteq \rho \vee \rho \nsubseteq \pi$
(vii) $\pi \nsubseteq \rho \vee \rho \nsubseteq \pi \supset \Box(\pi \nsubseteq \rho) \vee \Box(\rho \nsubseteq \pi)$ [(R$\subseteq$ ii), sentential logic]
(viii) $\Box(\pi \nsubseteq \rho) \vee \Box(\rho \nsubseteq \pi) \supset \Box(\pi \nsubseteq \rho \vee \rho \nsubseteq \pi)$ [distribution of $\Box$ over $\vee$]
(ix) $\pi \neq \rho \supset \Box(\pi \neq \rho)$ [(vi), (vii), (viii), (iv), K],

where the last step is obtained by substitution of equivalents, which is sanctioned by the principles of K.

MEMO TO AUTHORS

The Journal of Philosophy is published monthly, in issues that (ordinarily) contain only 56 pages of copy, this limits the length of papers the editors can normally publish to no more than 7,500 words, or 18 *Journal* pages A more precise way of determining length is by characters (including spaces) 2,850 characters per *Journal* page, or 51,300 characters per *Journal* article of 18 pages

The editors request that three copies (or an electronic version) of all manuscripts submitted be prepared according to the following guidelines

(1) All material, including the quotations and footnotes, must be typed, double-spaced, and one-sided, with ample margins left and right, top and bottom
(2) No bibliographies will be published, bibliographical material should be put into footnotes
(3) Footnotes should be complete, according to *Journal* style (see any issue) Journal citations should include name of journal in full, volume number, issue number, month or season, year, and pages Book references should include publisher, U S city of publication, and date Footnotes should be gathered at the end of the article
(4) Quotations should be checked by the author, capitalization, italics, spelling, and the like should match the original
(5) If there are diagrams, they should be provided camera-ready by the author If there are special symbols, they should be widely spaced and clearly marked Standard logic symbols are available, and letters of the Greek and script alphabet
(6) No underlining, please limit use of italics to technical terms (when they are introduced) and to variables
(7) No contractions

The editors are willing to read and evaluate a manuscript that does not meet these guidelines, provided the author agrees to meet them before submitting a final manuscript The editors may ask for revision of excessively long manuscripts, and may require reformatting of accepted manuscripts They also request an electronic version with the final hard copy

The Journal of Philosophy does not publish papers that are primarily historical or expository; the editors urge that such manuscripts be submitted elsewhere While we do not publish manuscripts that mainly comment on articles published elsewhere, we do have a "Comments and Criticism" section in which short responses to *Journal* articles may be published

Manuscripts may be sent to John Smylie, Managing Editor, *The Journal of Philosophy*, 709 Philosophy Hall, 1150 Amsterdam Avenue, Mail Code 4972, New York, NY 10027, Email jys3@columbia edu

NEW BOOKS ANTHOLOGIES

GRANT, RUTH W , ed *Naming Evil, Judging Evil* Chicago, University Press, 2006 xii, 234 p Cloth $35 00

GREEN, JOEL B , and STUART L PALMER, eds *In Search of the Soul Four Views of the Mind-Body Problem* Downers Grove, IL Intervarsity, 2005 223 p Paper $20 00

GREEN, RICHARD, and K SILEM MOHAMMAD, eds *The Undead and Philosophy Chicken Soup for the Soulless* Chicago Open Court, 2006 xvi, 261 p Paper $17 95

GREEN, MITCHELL, and JOHN N WILLIAMS, eds *Moore's Paradox New Essays on Belief, Rationality, and the First Person* New York Oxford, 2007 x, 247 p Cloth $65 00

GREENBERG, JEFF, SANDER L KOOLE, and TOM PYSZCZYNSKI, eds *Handbook of Experimental Existential Psychology* New York Guilford, 2004 xvi, 528 p Cloth $65 00

GREENOUGH, PATRICK, and MICHAEL P LYNCH, eds *Truth and Realism* New York Oxford, 2006 ix, 253 p Paper $35 00

GRECO, JOHN, ed *Ernest Sosa and His Critics* Malden, MA Blackwell, 2004 xxiv, 331 p Paper $34 95

GUINN, DAVID E , ed *Handbook of Bioethics* New York Oxford, 2006 xv, 437 p Cloth $99 00

GUTTING, GARY, ed *The Cambridge Companion to Foucault, 2nd Edition* New York Cambridge, 2005 xix, 465 p Cloth $70 00, paper $28 99

GUYER, PAUL, ed *The Cambridge Companion to Kant and Modern Philosophy* New York Cambridge, 2006 xv, 722 p Cloth $75 00, paper $29 99

HAACK, SUSAN, ed *Pragmatism, Old and New Selected Writings* Amherst, NY Prometheus, 2006 741 p Cloth $32 00

HAAKONSSEN, KNUD, ed *The Cambridge History of Eighteenth-Century Philosophy* New York Cambridge, 2006 1424 p in 2 volumes Cloth $250 00, *The Cambridge Companion to Adam Smith* New York Cambridge, 2006 xiv, 409 p Cloth $70 00, paper $28 99

HALWANI, RAJA, ed *Sex and Ethics Essays on Sexuality, Virtue, and the Good Life* New York Palgrave, 2007 xiv, 276 p Cloth $74 95

HAILA, YRJÖ, and CHUCK DYKE, eds *How Nature Speaks The Dynamics of the Human Ecological Condition* Durham, NC Duke, 2006 viii, 334 p Cloth $84 95, paper $23 95

HAMINGTON, MAURICE, and DOROTHY C MILLER, eds *Socializing Care Feminist Ethics and Public Issues* Lanham, MD Rowman and Littlefield, 2006 xxii, 233 p Cloth $75 00, paper $28 95

HANLEY, RICHARD, ed *South Park and Philosophy Bigger, Longer, and More Penetrating* Chicago Open Court, 2007 vii 302 p Paper $17 95

HARDCASTLE, GARY L , and GEORGE A REISCH, eds *Monty Python and Philosophy* Chicago Open Court, 2006 xi, 276 p Paper $17 95, *Bullshit and Philosophy* Chicago Open Court, 2006 xxiii, 272 p Paper $17 95

HART, KEVIN, ed *Counter-Experiences Reading Jean-Luc Marion* Notre Dame University Press, 2007 ix, 478 p Paper $40 00

New from OXFORD

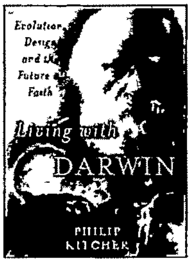
The Ethics of Care
Personal, Political, Global

VIRGINIA HELD

"Virginia Held's theory of care is no less substantial than John Rawls' theory of justice. Her probing and engaging analysis of caring values, virtues, actions, and attitudes will become a classic in moral theory."—Rosemarie Tong, The University of North Carolina at Charlotte

What is the ethics of care? Is it any better than other moral theories in enabling us to deal with the actual problems we confront? This book makes clear how the ethics of care differs from such dominant theories as Kantian moral theory, utilitarianism, and virtue ethics, and how promising it is as an approach to morality in any context.

2005 (paper February 2007) 224 pp paper $24.95 cloth $45.00



Living with Darwin
Evolution, Design, and the Future of Faith

PHILIP KITCHER

"Kitcher has just the combination of philosophical talent, biological insight, and wonderfully lucid writing needed to address the thorny problem of creationism. In *Living With Darwin*, he clearly shows that the persistent battle between evolution and creationism in America is part of a wider war—one between superstition and rationality. His analysis of this conflict, and suggestions for its resolution, should be read by everyone concerned with the relationship between faith and science."—Jerry Coyne, University of Chicago

March 2007 208 pp., 3 line illus. $20.00

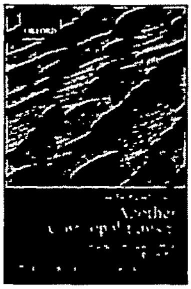
New in Paperback!

The Magic Prism
An Essay in the Philosophy of Language

HOWARD WETTSTEIN

In this accessible study, Wettstein argues that Wittgenstein, a figure with whom the critics of Frege and Russell are typically unsympathetic, laid the foundation for much of what is revolutionary in recent developments in the movement of philosophy of language.

2004 (paper September 2006) 256 pp paper $24.95 cloth $55.00



Another Cosmopolitanism

SEYLA BENHABIB

Edited by ROBERT POST

Benhabib argues that since the UN Declaration of Human Rights in 1948, we have entered a phase of global civil society which is governed by cosmopolitan norms of universal justice. The volume contains a substantive introduction by Robert Post, the volume editor, and contributions by panelists Bonnie Honig, Will Kymlicka, and Jeremy Waldron.
(The Berkeley Tanner Lectures)

2006 224 pp $25.00

OXFORD
UNIVERSITY PRESS

Prices are subject to change and apply only in the US. To order, or for more information, please call 1-800-451-7556. In Canada, call 1-800-387-8020. **Visit our website at www.oup.com/us**

In Memoriam

ROBERT C. SOLOMON

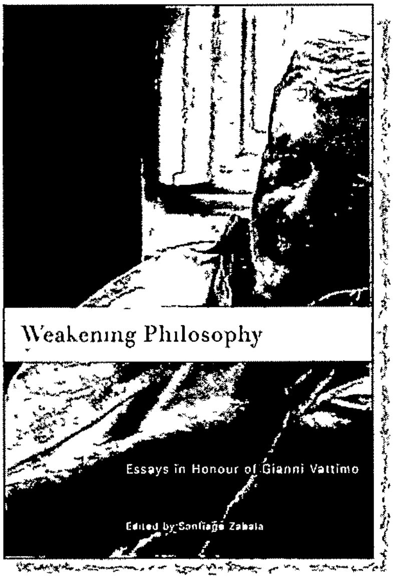
PHILOSOPHER

AUTHOR

FRIEND

1942 – 2007

OXFORD
UNIVERSITY PRESS



Weakening Philosophy

Essays in Honour of Gianni Vattimo

Edited by Santiago Zabala

978-0-7735-3143-7 $27.95 paper
978-0-7735-3142-0 $85.00 cloth

"*Weakening Philosophy* does what no other book has done with this major figure in contemporary thought: place him at the centre of a dialogue featuring the most stellar voices in contemporary philosophy – Eco, Nancy, Rorty, Taylor, Frank, Welsch, and Oñate."

William Egginton, romance languages and literatures, SUNY Buffalo

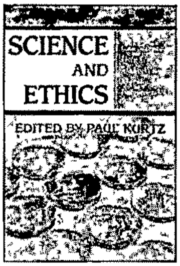
"An outstanding and unique collection. Vattimo has made numerous original, path-breaking contributions to the philosophy of religion, all of which are very intelligently discussed, analyzed, and criticized by the contributors."

Barry Allen, philosophy, McMaster University

McGILL-QUEEN'S UNIVERSITY PRESS

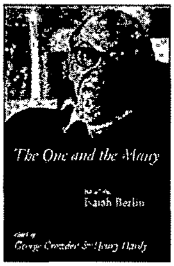
www.mqup.ca

SCIENCE AND ETHICS
Can Science Help Us Make Wise Moral Judgments?
Edited by Paul Kurtz, *with the assistance of David Koepsell*
350 pp • ISBN 978-1-59102-537-5 • PB • $21
A unique collection of authors who maintain that science can help us make wise choices and that an increase in scientific knowledge can help modify our ethical values and bring new ethical principles into social awareness.



PUTTING PHILOSOPHY TO WORK
Inquiry and Its Place in Culture—*Essays on Science, Religion, Law, Literature, and Life*
Susan Haack
336 pp • ISBN 978-1-59102-528-3 • HC • $32
This wide-ranging collection of Haack's new and recent essays offers applied philosophy in a new key and a fresh voice. Whether she is telling the "sad, sleazy saga" of a drug company's efforts to suppress the unfavorable results of a clinical trial it had funded, disentangling the many senses of "coherent," or reflecting on the growth of intellectual character, she writes with the clarity, verve, and wit her readers have come to expect.

THE ONE AND THE MANY
Reading Isaiah Berlin
Edited by George Crowder and Henry Hardy
300 pp • ISBN 978-1-59102-448-4 • HC • $28
"[A] fitting tribute to Berlin. It is a multidimensional exploration of the man and the ideas: his Judaism, his moderate nationalism, his ambivalence about the Enlightenment, his reaction to Stalinism, and a great deal more. It is a rich, original, and nourishing volume."
—Alan Ryan, warden of New College, Oxford

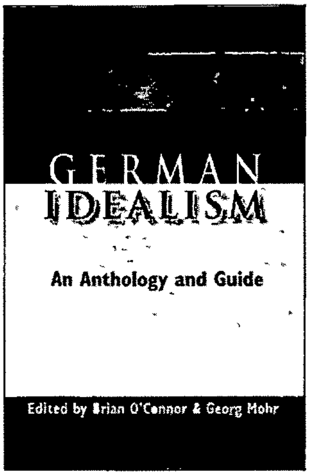


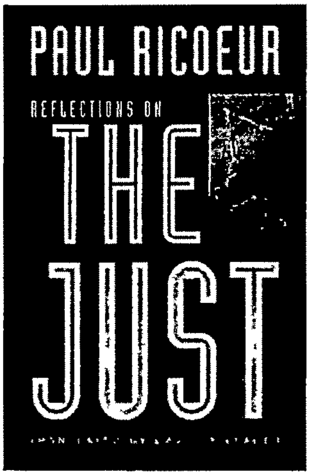
SKEPTICISM
An Anthology
Edited by Richard H. Popkin and José R. Maia Neto
350 pp • ISBN 978-1-59102-474-3 • PB • $29
This anthology contains the principal texts of the Western skeptical tradition from its origins in antiquity to contemporary philosophy. An introduction on the origins, kinds, and significance of philosophical skepticism puts the various readings into historical context.

DENNETT AND RICOEUR ON THE NARRATIVE SELF
Joan McCarthy
220 pp • ISBN 978-1-59102-548-1 • PB • $39.95
Is the self some kind of real entity or a very convincing illusion? Drawing on the work of two leading philosophers, Daniel Dennett and Paul Ricoeur, McCarthy examines how each of these thinkers casts the self in narrative terms.

Prometheus Books
Toll Free 800-421-0351 / Fax 716-691-0137
www.prometheusbooks.com

Philosophy *New from Chicago*





German Idealism
An Anthology and Guide
Edited by Brian O'Connor and Georg Mohr

"This is a welcome and helpful collection of key texts by the four great German Idealists which will be a very valuable resource for students and teachers alike."—Stephen Houlgate, University of Warwick
Paper $37.50

Reflections on the Just
Paul Ricoeur
Translated by David Pellauer

Consisting of fifteen thematically organized essays, *Reflections on the Just* continues and expands on the work Ricoeur began with his "little ethics" in *Oneself as Another* and *The Just*. Cogent, deeply considered, and fully engaged with the realities of the contemporary world, *Reflections on the Just* is an essential work for understanding the development of Ricoeur's thought in his final years.
Cloth $25.00

Now in paperback

Hope Now
The 1980 Interviews
Jean-Paul Sartre and Benny Lévy *Translated by Adrian van den Hoven*

"Just as Sartre had at first preached existentialism, then Marxism, here Sartre expresses belief in a political hope, replacing Marxism with ethics."—*Library Journal*
Paper $15.00

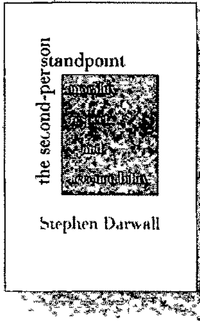
The University of Chicago Press • www.press.uchicago.edu

RADICAL HOPE
ETHICS IN THE FACE OF CULTURAL DEVASTATION
JONATHAN LEAR

"With admirable lucidity, in the most clear-cut language, he shows us that besides the glamorous alternatives of freedom or death there is a third way, less grand yet demanding just as much courage the way of creative adaptation"
—J M COETZEE
Smith Fund • New in Cloth

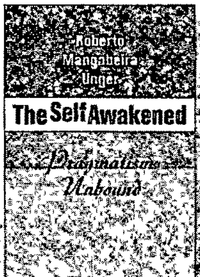
THE SECOND-PERSON STANDPOINT
MORALITY, RESPECT, AND ACCOUNTABILITY
STEPHEN DARWALL

Darwall elaborates the interpersonal nature of moral obligations their inherent link to our responsibilities to one another as members of the moral community
New in Cloth



THE SELF AWAKENED
PRAGMATISM UNBOUND
ROBERTO MANGABEIRA UNGER

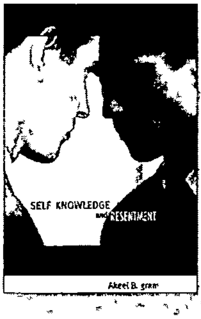
The Self Awakened mobilizes the resources of several philosophical traditions, and develops the unrecognized revolutionary implications of the most influential of these traditions today—pragmatism
New in Cloth



WWW.HUP.HARVARD.EDU HARVARD UNIVERSITY PRESS

SELF-KNOWLEDGE AND RESENTMENT

AKEEL BILGRAMI



"[A] distinctive and original treatment which covers an impressive sweep of philosophical ground and creates a whole new framework for thinking about self-knowledge and the philosophical landscape around it"

—CRISPIN WRIGHT, UNIVERSITY OF ST ANDREWS

New in Cloth

THE VEIL OF ISIS

AN ESSAY ON THE HISTORY OF THE IDEA OF NATURE

PIERRE HADOT

Translated by Michael Chase

"The book is a dizzying *tour de force* that would be the envy of a modern Plotinus"

—G W BOWERSOCK

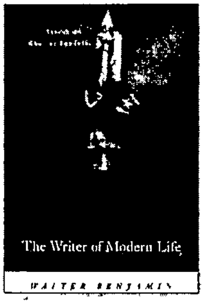
Belknap Press • New in Cloth

THE WRITER OF MODERN LIFE

ESSAYS ON CHARLES BAUDELAIRE

WALTER BENJAMIN

Edited by Michael W. Jennings
Translated by Howard Eiland, Edmund Jephcott, Rodney Livingstone, and Harry Zohn



"In these essays, written in the 1930s, German critic Benjamin masterfully succeeds in changing our perception of French poet Charles Baudelaire as a late Romantic dreamer Stimulating reading"

—BOB T IVEY, *LIBRARY JOURNAL*

Belknap Press • New in Paperback

WWW.HUP.HARVARD.EDU HARVARD UNIVERSITY PRESS

ANNOUNCING:

THE JOURNAL OF PHILOSOPHY

ONLINE

The Journal of Philosophy has added
a print and online subscription option for institutions

SUBSCRIPTION OPTIONS

Libraries and Institutions (print and online)	$180.00
Libraries and Institutions (print only)	$100.00
Individuals	$45.00
Students, retired/unemployed philosophers	$20.00
Postage outside the U.S.	$15.00

Please consult our website at www.journalofphilosophy.org
for more details

Calcutta University Library
30.8.07
ALIPORE
CAMPUS

THE JOURNAL OF PHILOSOPHY

VOLUME CIV, NUMBER 2
FEBRUARY 2007

page
55 *Normativity and the Rejection of Rationalism* Lionel K. McPherson
71 *Presuppositions and Scope* Daniel Rothschild
107 NEW BOOKS ANTHOLOGIES

Published by **The Journal of Philosophy, Inc.**

THE JOURNAL OF PHILOSOPHY

FOUNDED BY FREDERICK J E WOODBRIDGE AND WENDELL T BUSH

Purpose To publish philosophical articles of current interest and encourage the interchange of ideas, especially the exploration of the borderline between philosophy and other disciplines

Editors Bernard Berofsky, Akeel Bilgrami, John Collins, Arthur C Danto, Kent Greenawalt, Patricia Kitcher, Philip Kitcher, Isaac Levi, Wolfgang Mann, Mary Mothersill, Frederick Neuhouser, Christopher Peacocke, Carol Rovane, Achille C Varzi *Consulting Editors* David Albert, James T Higginbotham, Robert May, Charles D Parsons, Wilfried Sieg *Managing Editor* John Smylie

THE JOURNAL OF PHILOSOPHY is owned and published by the Journal of Philosophy, Inc *President*, Arthur C Danto, *Vice President*, Akeel Bilgrami, *Secretary*, Daniel Shapiro, *Treasurer*, Barbara Gimbel, *Other Trustees* Lee Bollinger, Leigh S Cauman, Kent Greenawalt, Michael J Mooney, Lynn Nesbit

All communications to the Editors and Trustees and all manuscripts may be sent to John Smylie, Managing Editor, Mail Code 4972, 1150 Amsterdam Avenue, Columbia University, New York, New York 10027 FAX (212) 932-3721

You may also visit our website at www journalofphilosophy org

THE JOURNAL OF PHILOSOPHY

2007

SUBSCRIPTIONS (12 issues)

Individuals	$45 00
Libraries and Institutions	$100 00
Libraries and Institutions (print and online)	$180 00
Students, retired/unemployed philosophers	$20 00
Postage outside the U S	$15 00

Payments only in U S currency on a U S bank All back volumes and separate issues available back to 1904 Please inquire for advertising rates, ad space is limited, so ad reservations are required

Published monthly as of January 1977, typeset by Dartmouth Journal Services, Waterbury, VT, and printed by The Sheridan Press, Hanover, PA

All communication about subscriptions and advertisements may be sent to Pamela Ward, Business Manager, Mail Code 4972, 1150 Amsterdam Avenue, Columbia University, New York, NY 10027 (212) 866-1742

All materials published in *The Journal of Philosophy* are copyrighted by the *Journal* and are protected by U S copyright law All rights reserved For more information about our copyright policies, please consult our January issue or contact us

The Journal of Philosophy (ISSN 0022-362X) is published monthly by The Journal of Philosophy, Inc , MC4972, Columbia University, 1150 Amsterdam Avenue, New York, NY 10027 Periodicals postage paid at New York, NY, and other mailing offices

POSTMASTER Please send address changes to *The Journal of Philosophy* at MC4972, Columbia University, 1150 Amsterdam Avenue, New York, NY 10027

© Copyright 2007 by the Journal of Philosophy, Inc

ISSN 0022–362X

University
ALIPORE
CAMPUS

THE JOURNAL OF PHILOSOPHY
VOLUME CIV, NO 2, FEBRUARY 2007

NORMATIVITY AND THE REJECTION OF RATIONALISM*

A normative reason is a consideration that supports a certain action, belief, or attitude While normative reasons usually are contrasted with explanatory reasons—why a person should φ versus why a person does φ—normativity has seemed puzzling [1] It is not obvious whether normative reasons in general, on their own merits, can compel us to φ Nor is it clear how widely normative reasons apply in similar circumstances, to all persons or perhaps only to some These issues arise even on a conception of persons as rational agents

Morality is often thought to provide normative reasons that doing *x* is morally wrong would be a reason not to do *x* If we are rationally required to do what morality requires, the "binding" nature of morality could not depend on any contingent feature of a rational agent's psychology Yet there are persons—call them *amoralists*—who believe they are rationally entitled to ignore the claims that morality makes To amoralists, moral considerations seem to provide no more reason for action than nonmoral interests provide If there is contingency in what persons can have reason to do, despite morality's claims, amoralists would not necessarily be irrational Whether particular persons have reason to do what morality requires would depend on what they happen to care about [2]

T M Scanlon has proposed a challenging account of moral normativity that some commentators believe shares a commitment to the

---

* I am grateful to David Copp, Erin Kelly, Doug Lavin, and Angela Smith for comments and discussion that significantly shaped the final version of this paper

[1] My topic is normative reasons So the "reasons" I refer to will be normative reasons, not explanatory reasons and not whatever a person might take to be normative reasons

[2] For critical discussion of the view that the contingency of moral values would undermine our commitment to them, see James Lenman, "Michael Smith and the Daleks Reason, Morality, and Contingency," *Utilitas*, xi (1999) 164–77

0022-362X/07/0402/55–70

rationalist view that “motivation to act in accord with reasons is entailed by rationality.”[3] In fact, Scanlon’s account is ambiguous: its loyalties seem divided between the rationalist notion that all persons have moral reasons, which by themselves motivate insofar as persons are acting rationally, and the rejection of rationalism. I explore this tension and defend a resolution that fits with my doubts about the plausibility of a rationalist account of normativity. Specifically, I argue that there can be reason to do what morality requires, where this does not depend on any contingent feature of a particular person’s psychology—without this entailing that everyone is rationally bound by morality’s claims or at least has moral reasons.[4] The rejection of rationalism does not threaten the normativity of morality. This normativity comes from the bona fide reasons that morality provides. Furthermore, practical features of moral life—which include reactive attitudes, moral criticism, and punishment—are viable even if all rational persons do not have moral reasons. The problem of normativity is not as formidable as the contemporary philosophical debate would lead us to believe.

I

A good place to start is by stepping back from an approach that has us think in terms of *the* question of normativity. Such an approach is found in Christine Korsgaard’s *The Sources of Normativity*.[5] “When we seek a philosophical foundation for morality,” Korsgaard writes, “we are not looking merely for an explanation of moral practices. We are asking what *justifies* the claims that morality makes on us. This is what I am calling ‘the normative question’” (*ibid.*, pp. 9–10).

“The normative question” is not easy to pin down, and much of the difficulty can be traced to slippage between agential perspectives. To ask, as Korsgaard does, “Why should I be moral?” seems straightforward (*ibid.*, p. 9). This question points toward an account of the reasons “I” have, if any, without obviously addressing how widely shared these reasons are. Korsgaard also asks how it could be that moral standards "make *claims* on us" and that "when we invoke them, we make claims on one another" (*ibid*, p 8) Of course, "we" often refers to a collection of individuals under some description (for example, rational beings), where there is no assumption or requirement that they share certain values or commitments But Korsgaard's "we" is stronger she collectivizes persons in a way that represents a distinctive, first-person perspective The implications of such a move are critical She cites Kant's view that "when we judge something beautiful we not only take pleasure in it, but demand that everyone do so" (*ibid*, p 8, note 8) [6] Translating this thought from aesthetics to morality, when I judge that an action is right, I would be demanding that everyone do it (*ibid*, p 8) Korsgaard thus seems to go further than raising the question "Why should I be moral?" such that if an answer can be given from one's own perspective as a rational being, then any other rational being is in a position to answer likewise The idea is not simply that the answer is available to all rational beings, for this would not warrant talk of a demand that everyone share one's own judgment and act accordingly Rather, Korsgaard offers an account of rationality in relation to morality that would resolve the ostensibly different questions of normativity she raises "Why should I be moral?" and "Why should everyone be moral?" become versions of the same question

---

[3] The quote is from David Copp, “Against Internalism about Reasons—Gert’s Rational Options,” *Philosophy and Phenomenological Research*, LXII (2001): 455–61, at p. 455. In this essay, I will focus on and refer to Scanlon’s *What We Owe to Each Other* (Cambridge: Harvard, 1998).

[4] I do assume that persons who accept that they have reason to act morally are disposed to act accordingly. This might make me an “internalist,” at least on some understandings of internalism. See, for example, Michael Smith, *The Moral Problem* (Cambridge: Blackwell, 1994). As Smith roughly describes internalism, “*Believing I should* [do what morality requires] seems to bring with it *my being motivated to*—at least absent weakness of the will and the like” (p. 60).

[5] Korsgaard, *The Sources of Normativity* (New York: Cambridge, 1996), pp. 9–10.

In short, "the normative question" can depend for its meaning on the presuppositions that may be attached to "I" and "we" Korsgaard shifts from the first-person perspective of a rational agent, who must value his own humanity, to a universal perspective of persons who, by virtue of their rationality and common humanity, must value the humanity of one another This is supposed to show that certain actions are obligatory for all rational beings, and that persons who fail to recognize this and to act accordingly are irrational due to inconsistency [7] My proposal is that the "I" and "we" questions do not necessarily yield a unified answer

We can proceed by roughly distinguishing two questions that correspond to the first-person and the universal perspectives

(1) Do I have reason to do what morality requires?
(2) Does every person have reason to do what morality requires?

Although I may seem to have reason to act morally, whether all rational persons have reason would remain to be argued Persons who acknowledge no commitment to morality feel that its claims lack normative force for them. An answer to the second question would therefore need grounding in a source of obligation that is independent of a person's commitments (insofar as those commitments do not come merely, perhaps unwittingly, from the fact of rational agency). Morality's claims, whether they motivate all rational persons, could then be binding on all rational persons. Some moral skeptics deny that an independent source of obligation can be identified, and they think that no one has moral reasons without it. Moral philosophers who are baited by this skepticism respond by trying to show that everyone has reason to act morally.

---

[6] Korsgaard references Kant's *Critique of Judgment*, part I, division I, book I, "The Analytic of the Beautiful"

[7] For critical discussion of Korsgaard's argument, see G A Cohen, "Reason, Humanity, and the Moral Law," in *The Sources of Normativity*, especially pp 172–77

Despite the appeal of answering the second question of normativity, answering the first is more important. The first question poses the troubling skeptical threat. If we can credibly answer this question, the second should not haunt us theoretically. Those of us who are not disinterested in morality believe that a person can have reason to act morally, even when morality conflicts with that person's nonmoral interests. We can resist the first-person skeptical challenge, since the question it raises can be addressed by appreciating why morality can prove compelling. Kant gives an answer that seems simple: that an action is right is a reason for a person to do it, given that ostensible moral reasons dictating rightness are bona fide reasons for action. We do not have to be Kantians about the content of morality to accept Kant's answer to the first question. But we do need some understanding of why moral reasons are bona fide reasons, we do not simply assume that there is reason to care about doing "the right thing." Call this the *dogmatism worry*. We are looking for bona fide moral reasons, as compared to whatever considerations someone might think are moral reasons (for example, that a religious figure has issued decrees), because the former are the basis for taking morality seriously.

Bona fide reasons are not restricted to moral reasons. Prudential reasons, for example, count as bona fide reasons as well. Moral reasons will reflect considerations that morality is characteristically responsive to, including respect for persons and concern for their welfare. To fill out this idea, Thomas Nagel's defense of altruistic reasons can be generalized and modified: the circumstances that moral reasons would have us care about regarding the lives of others are circumstances that those others already have reason to care about from a self-interested point of view.[8] There is disagreement, no doubt, about how exactly to specify such circumstances. Nevertheless, they broadly enter into an explanation of why moral reasons are bona fide reasons This shows how the substance of moral argument can help to break the spell that moral skepticism may have cast on persons otherwise disposed to care about morality The target audience is persons who believe that morality has some authority but are in doubt about the content of moral requirements and how much weight they carry Moral argument cannot reach persons who are determined to believe that morality is for suckers or is of no special importance relative to their nonmoral priorities

---

[8] See Nagel, *The Possibility of Altruism* (Princeton University Press, 1970), p. 16. I am not endorsing Nagel's "general thesis" that "one has a *direct* reason to promote the

More than a few philosophers will remain troubled Ultimately, they seem to have in mind the second question of normativity and its universal perspective They do not want to know merely why I or you, from a first-person perspective, may have reason to act morally They want to know, further, what may justify the demand that everyone do what morality requires In the background is often the assumption that morality is so important that we cannot opt out if we are rational, that the legitimacy of morality's claims would otherwise be undermined I am arguing that the assumption is gratuitous a credible account of moral normativity is no worse off without it

II

The failure of some people to take morality seriously should not drive philosophers to exasperation Kantians, for example, charge amoralists with irrationality Yet amoralists might well be capable of leading functional lives and accommodating themselves to the reality that other people take morality quite seriously *Strong rationalism,* as I will refer to it—the view that we are rationally required to do what morality requires and, hence, that we are irrational in failing to do so—seems either descriptively false or tendentiously true Indifference to morality's claims is distinctively and most straightforwardly a moral failure demonstrated, for instance, by willingness to free ride on the good will of others The putative inconsistency displayed by free-riding amoralists who would value their own humanity and self-interestedly hope that it is valued by others, without generally valuing the humanity of others, is not inherently at odds with the prudential pursuit of their own aims

Unlike strong rationalists, Scanlon is prepared to refrain from charging amoralists with irrationality He accepts that we are not rationally

---

interests of others—a reason which does not depend on intermediate factors such as one's own interests or one's antecedent sentiments of sympathy and benevolence" (pp 15–16) required to grant morality importance: while the kinds of considerations that morality is characteristically responsive to may not resonate with some persons, a failure of rationality cannot necessarily be attributed to such persons, though they may be lacking in other, significant respects. At the same time, Scanlon rejects strong rationalism en route to a different type of rationalism: he believes that a person's sense of moral obligation and her accompanying moral motivation can derive solely from her acknowledgment of moral reasons, reasons that all rational persons have. I will argue that Scanlon's approach to normativity would be more convincing and less misleading if he rejected rationalism generally—that is, any view according to which all rational persons have and can be directly motivated by moral reasons. We need not accept rationalism in order to take the idea of moral reasons seriously. Nor are our moral practices threatened in denying that a person's moral motivation can be explained solely by appealing to her acknowledgment of moral reasons.

A *nonrationalist* approach to normativity, as I will call it, can maintain that morality provides reasons rather than only expresses pro or con attitudes. Mill suggests such an account. What he calls "the internal sanction of duty" is "a feeling in our own mind, a pain ... attendant on violation of duty, which in properly cultivated moral nature rises ... into shrinking from it as an impossibility. This feeling, when disinterested and connecting itself with the pure idea of duty, and not with some particular form of it, or with any of the merely accessory circumstances, is the essence of conscience ...."[9] Implicit here is the view that for those of us who possess a moral sensibility, the reason we have not to violate duty is that doing so would be wrong. This bears a striking similarity, mutatis mutandis, to Kant's view.[10] Mill goes on to explain how people may come to care about morality and have the feeling of conscience: the force of morality's claims resonates through "the desire to be in unity with our fellow creatures, which is already a powerful principle in human nature ..." (*ibid.*, p. 32). He does not argue that everyone must have this desire. Persons who are motivated by morality recognize reasons that can roughly be described as agent-neutral but not universal.[11] These reasons would yield an answer to the first question of normativity: I have reason to do what morality requires because moral reasons are reasons that persons who have the desire to be in unity with others have, and I am such a person "Undoubtedly this sanction has no binding efficacy," Mill writes, "on those who do not possess the feelings it appeals to On them morality of any kind has no hold but through the external sanctions" (*op cit*, p 29) Thus the answer to the second question of normativity would be no everyone might not have reason to do what morality requires

---

[9] Mill, *Utilitarianism*, George Sher, ed. (Indianapolis: Hackett, 1979, 2nd ed.), pp. 28–29.

[10] Mill does emphasize that conscientious persons experience the wrongness of violating duty as painful. This does not mean, however, that the wrongness consists in or originates in the violation disturbing our conscience.

[11] My reading of Mill might seem idiosyncratic. But a similar approach to moral normativity can be found in other utilitarians. J. J. C. Smart, for example, observes that

Scanlon cites with approval the connection Mill makes between moral duty and moral motivation This connection bridges the gap between having reason to act morally and wanting to act morally "What we need to do, then," Scanlon writes, "is to explain more clearly how the idea that an act is wrong flows from the idea that there is an objection of a certain kind to people's being allowed to perform such actions, and we need to do this in a way that makes clear how an act's being wrong in the sense described can provide a reason not to do it" (*op cit*, p 153) In particular, Scanlon observes that his contractualist idea of wanting to be able to justify our actions to others is similar to Mill's idea of the desire to be in social unity But Scanlon rejects Mill's account of moral motivation, which is to say that Scanlon rejects Mill on the internal sanction of morality "[T]here is no need to appeal to a special psychological element to explain how a person could be moved to avoid an action by the thought that any principle allowing it would be one that others could reasonably reject This is adequately explained by the fact that people have reason to want to act in ways that could be justified to others, together with the fact that when a rational person recognizes something as a reason we do not need a further explanation of how he or she could be moved to act on it" (*op cit*, p 154)

Part of Scanlon's criticism of Mill is fair Mill's appeal to the desire to be in social unity does not affirm that morally salient considerations support having the desire Although he tells a story about how we may come to have the desire to be in social unity, he does not state that the desire is an appropriate source of motivation for a rational person However, Mill could accept that morally salient considerations support having the desire, he is not tied to brute desire in explaining moral motivation His focus is on describing an ordinary process of socialization that leads, in normal cases, to persons acquiring a moral sensibility Mill does hold that reasons by themselves cannot motivate, and

---

utilitarianism is supposed to apply to "sympathetic and benevolent" persons, see Smart and Bernard Williams, *Utilitarianism For and Against* (New York Cambridge, 1973), p 31 Also see Henry Sidgwick in *The Methods of Ethics* (Indianapolis Hackett, 1981) this would include what Scanlon identifies as reason to want to be able to justify your actions to others. you would have to care in the first place about such an aim. Now, Scanlon agrees that reasons by themselves cannot motivate. He emphasizes that reasons can motivate only when they are *acknowledged* by a rational agent. But this does not establish his distinctive claim that a rational person who acknowledges moral reasons will be motivated to act on them (barring, say, weakness of will). In general, for a rational person to acknowledge certain reasons is not necessarily for her to endorse them on grounds that would have everyone, including her, who acknowledges the reasons be guided by them.

Consider origami, which I recognize is an art. Because art is worth valuing, there is reason for me to care about origami. Yet my resources are limited. There are many more things that there is reason to care about than I can pursue. Human beings, with our finite lives and capacities, must have priorities. Scanlon does allow that some reasons have "subjective conditions," which leave room for variation in how persons respond to such reasons.[12] He also believes that if something is worthwhile, we have reason to have a positive attitude toward it. While I would not have reason to become an origami connoisseur, I should acknowledge a reason that motivates me at least to adopt an attitude of respect toward origami and those persons who do particularly care about it. The problem is this falls short of the more robust sense in which Scanlon thinks that all rational persons have moral reasons. He does not believe that normal variation in our responses is so wide as to permit a rational person somehow to respect moral reasons without being motivated to act on them.

Instead, Scanlon believes that morality is special. The binding nature of moral reasons for a rational person who acknowledges them, he maintains, is captured by "the great importance of justifiability to others and to the particular interests that moral principles protect," and also by the idea that "the other [nonmoral] values, properly understood, have a built-in sensitivity to the demands of right and wrong" (*op. cit.*, p. 166). Morality, in contrast to things like origami, structures fundamental aspects of our interactions with other people. If you acknowledge moral reasons, you cannot treat them as optional. Yet the argument that moral reasons, being special in this sense, can directly motivate us does not flow from a general claim about the motivation that a reason can supply by itself for rational persons who acknowledge it. Rather, the argument flows from Scanlon's more specific claim about the priority of moral reasons and their accompanying motivational force for *reasonable* persons, that is, for persons who care about being able to justify their actions to one another This leaves open the possibility that some rational persons could acknowledge that there are moral reasons without granting these reasons priority

---

[12] See, for example, Scanlon, p. 370.

The point of contention here is why Scanlon seems to discount this possibility by arguing that when rational persons acknowledge moral reasons, no further explanation is needed of their motivation to act accordingly My diagnosis is the following Moral reasons are supposed to be available to every rational person Persons who acknowledge these reasons are rationally required to act as morality requires, since moral reasons in particular warrant acknowledgment as reasons of a special type that would obligate and motivate all rational persons But Scanlon is not entirely comfortable with this position He realizes that not every rational person will be motivated by morality Like Mill, he accepts that morality can provide no reasons that motivate for persons who lack the kind of concern for others that moral reasons reflect More surprisingly, Scanlon also accepts that persons who are not motivated by morality are not necessarily irrational [13] This sets him apart from strong rationalists Then, again, Scanlon believes that moral reasons "apply" to amoralists as well What could this mean? At issue is the significance of the possibility of a rational person acknowledging moral reasons without granting them priority

III

I have shown how Scanlon articulates a position that stops short of strong rationalism He argues that when rational persons take themselves to have moral reasons, their motivation to act accordingly can be *explained* by that fact This is weaker than the strong rationalist claim that moral motivation is *entailed* by rationality The explanatory claim allows that not all rational persons must take themselves to have moral reasons [14] Some commentators have missed this difference between the entailment claim made by strong rationalists and the explanatory claim Scanlon makes [15] Yet Scanlon seems to endorse a version of rationalism when he describes the situation of amoralists. Like the rest of us, they are supposed to have moral reasons, which could directly motivate them: “we must say that the moral reasons that apply to us apply to these people as well. . . . Failure to care about right and wrong does not make a person irrational . . . but a person who is left cold by moral considerations does fail to appreciate reasons that apply to him or her” (*op. cit.*, p. 158). Gerald Dworkin finds such a stance troubling, for this sense in which moral principles would apply to rational persons seems to him too weak to account for the ordinary, strong sense in which moral principles would apply.[16]

---

[13] This is less of a concession than it might seem Scanlon characterizes amoralists as mistaken or "deficient" if they fail to recognize moral reasons and thus fail to act on them (p 370)

[14] Smith characterizes Scanlon's position as "externalist" since it rejects "the claim that there is a necessary connection of sorts between moral judgement and motivation" (*The Moral Problem*, p 12) But Scanlon might accept that there is a connection "of sorts," namely, when persons are reasonable

[15] See, for example, Copp, pp 457–58

When a moral principle is said to apply to rational persons, this suggests that a person who is not motivated by that principle is irrational. By contrast, as Dworkin interprets Scanlon, a moral principle *F* may be binding on amoralists, or the morally “disaffected,” in that “the principles that specify right and wrong apply to them so that if *F* says that someone acts wrongly in doing *x*, then a disaffected person acts wrongly in doing *x*” (*ibid.*, p. 478). The principle is binding from the perspective of moral judgment—a perspective that does not depend on the claim that a person who has acted wrongly is thereby irrational. This fits with Scanlon’s view that the case of the disaffected is captured by “the universality of reason judgments”: “their case is quite different from that of people who ‘have different tastes’. . . . In these cases, the main point of the activities in question is a certain kind of enjoyment, so people who do not get this enjoyment from the activities lack reasons to engage in them. But morality is not aimed at enjoyment, so the reasons to give it a place in one’s life are not conditional in this way” (*op. cit.*, p. 158). Scanlon thinks that the disaffected can be criticized, even though they are not necessarily irrational in failing to be motivated by morality.

Reasonable persons, Scanlon contends, care about the justifiability of their actions to others. A person may be rational, though, and lack this concern. The gap between the rational and the reasonable is possible given a conception of irrationality that “fits better with ordinary usage and has other important advantages”: on Scanlon’s narrow construal, “Irrationality in the clearest sense occurs when a person’s attitudes fail to conform to his or her own judgments . . .” (*op. cit.*, p. 25). Nonetheless, Scanlon believes that persons who do not care about the justifiability of their actions to others do have moral reasons. The universality of reason judgments would imply more than that we can say of a rational yet disaffected person that she acts wrongly We also can say that she has moral reasons she does not acknowledge Every rational person has moral reasons, and this is not conditional on wanting to be in a certain kind of relationship or on any other contingent psychological feature

---

[16] Dworkin, “Contractualism and the Normativity of Principles,” *Ethics*, cxii (2002): 471–82.

Scanlon's account of moral normativity is confusing because he insists on a view—namely, all rational persons have and can be directly motivated by moral reasons—that appears to tie him to strong rationalism He seems to imply that insofar as a person is rational, she will be motivated by moral reasons But, as we have seen, Scanlon is not committed to this position He allows that a rational person might not acknowledge moral reasons, which could explain why some rational persons lack moral motivation (He also allows that some persons might acknowledge moral reasons and not be morally motivated, which he considers a case of the narrow irrationality of failing to do what one takes oneself to have conclusive reason to do) Moral reasons are sufficient by themselves to motivate only those rational persons who are reasonable in that they acknowledge they have reason to want to live with others on mutually justifiable terms

So Scanlon is not claiming that we can expect the disaffected, as rational persons, to be motivated by morality For Dworkin, this is a basis for the following objection "[If] the disaffected have no reason to act morally (because they do not share our ideal of reaching general agreement, and could not be brought to share it by any mode of reasoning or further experience), then we cannot say that such a person has a reason to not act immorally, only that they are deficient in not seeing that there is such a reason" (*op cit*, p 480) I am doubtful that this poses a genuine difficulty To illustrate why, I propose drawing a distinction between the idea that there are reasons to φ and the idea that a particular person has reasons to φ This helps to avoid confusion about perspective in talk of reasons, a problem Bernard Williams has highlighted [17] As Scanlon succinctly describes Williams's view, "when we speak of the reasons another person has, we are taking that person's 'perspective on the situation' and pointing out what is so, given that perspective" (*op cit*, p 372) The distinction between reasons there are to φ and a particular person having reasons to φ makes perspicuous the perspective of the person in question, without committing us to any specific view about the nature of reasons or what reasons the person should acknowledge Scanlon does not employ such a distinction, for he seems concerned that it obscures the universality of reason judgments. Yet his position overall would be better served if he were to respond to Dworkin that indeed the most we can claim—strictly regarding the relation between rational agency and reasons—is that there is reason for the disaffected not to act immorally. The weak sense in which the disaffected might acknowledge moral reasons, namely, as considerations that could motivate a rational person, does not support the claim that the disaffected must act morally if they are rational. They could act rationally if they instead pursue other worthwhile endeavors, despite a resulting conflict with morality.

---

[17] Williams, "Internal Reasons and the Obscurity of Blame," in *Making Sense of Humanity* (New York Cambridge, 1995), pp 35–45, see p 36

On the position I am defending, we can hope and expect that people develop the kind of sensibility that would render them receptive to acknowledging and being motivated by moral reasons. But we can give up on the claim that moral reasons must be universally normative: moral reasons might not "apply" to every rational person, except in the sense that, qua reasons, they are available to rational persons. Scanlon is correct that the disaffected are not necessarily irrational. Nothing is added, however, by the idea that moral reasons are "binding" for a rational person who acknowledges them: this idea merely marks the claim that a person who acts immorally acts wrongly. The disaffected fail to do what there is moral reason to do. Yet such a failure, at the level of the relation between rational agency and reasons, is akin to failing to care about origami, in the sense that one would be unwilling to forgo one's personal interests out of respect for it. Of course, the stakes are much greater in the case of morality. I am arguing, though, that the explanatory claim about moral motivation must presuppose that rational persons who are motivated by morality have a kind of concern for others that grounds their special sensitivity to moral reasons.

Intractable amoralists, by hypothesis, cannot be compelled to accept that they have bona fide reasons to act morally. That they lack a moral sensibility explains why they are not morally motivated. Obviously, this may have practical consequences for how they decide what to do and for how we treat them in response. But the practical consequences do not impugn the status of moral reasons as bona fide reasons that reasonable persons will have. If we resort to calling amoralists irrational, we are in fact representing our judgment about a broader human deficiency that leads them to fail to recognize or be motivated by the reasons there are to do what morality requires. The charge of irrationality will have no greater leverage against amoralists than the distinctively moral judgments we are already able to make about them. Nor does the charge illuminate some inherent inconsistency given what rational persons might value. Dworkin objects that if a particular person's having moral reasons is contingent on her having a moral sensibility, "then the authority of such principles for those who do not have such reasons is put more centrally into doubt" (*op cit*, p 482) Yet such a notion of authority is obscure whatever authority moral reasons have, they will be impotent, in effect, regarding amoralists

Let me be explicit about two respects in which my account of moral normativity is nonrationalist (1) it denies that all rational persons as such have reason to do what morality requires, (2) it denies that moral motivation can derive solely from a rational person's acknowledgment of moral reasons I am not endorsing the sweeping, Humean view that reason is a slave to the passions Rather, to reiterate, I have argued that in order to explain how persons could be morally motivated, we must presuppose that they have a kind of concern for others that renders them receptive to morality's claims This can make sense of the possibility that some persons, acknowledging moral reasons and acting rationally, might not grant priority to moral reasons For persons who do have a moral sensibility, moral reasons may appear to influence directly what they take themselves to have reason to do they can come to care about doing the right thing for the bona fide reasons they recognize make it right Such persons, we might say, are reasonable However, pace Scanlon, this idea of reasonableness does import a "psychological element" into the explanation of moral motivation for rational persons

IV

Rationalism—whether of the entailment-claim variety (for example, Kant's) or the explanatory-claim variety (for example, Scanlon's)—does not seem plausible There can be contingency in whether a rational person has special reason to value morality and, hence, in what a rational person has reason to do Nevertheless, along the lines of Mill and Scanlon, I have maintained that reasonable persons will have special reason to value morality and to act accordingly A person's disinterest in morality represents a human deficiency broader than irrationality she might not feel empathy or compassion, or might not have the desire to be in unity with other persons, or might not care about the justifiability of her actions to them Simply put, amoralists lack a basic regard for the lives of other beings, which is the mark of an essentially antisocial character But disinterest in morality does not involve a mistake of reason Or, at least, failure to care about morality is no more a mistake of reason than failure to care about art or any number of other endeavors that are worthwhile

My defense of nonrationalism returns us to the moralist's worry about the threat to practical features of moral life I want to examine what triggers the notion that rationalism had better be true—that, otherwise, our commitment to morality would be undermined. As Dworkin expresses the worry over amoralists, "The very thing which explains our motivation to act rightly has no motivational force with them, and cannot, given their (defective) appreciation of the value of being in a certain relationship with others. But then not only does [moral] criticism lose its point, so does a whole range of what Strawson calls reactive attitudes" (*op. cit.*, p. 481). Similarly, Michael Smith believes that if we stopped caring or had never cared about morality, we would "rightly come to see" our caring about it "as *rationally optional,* as "*arbitrary,* as to be *explained* rather than *justified* . . . simply in terms of *when, where* and *how we were brought up*."[18]

We can ask whether the dire conclusions that Smith and Dworkin draw would follow. After all, reasonable persons can take morality seriously even if some persons do not. The possibility that there could be rational amoralists need provide no impetus for the rest of us to lose interest in morality's claims. As long as enough people share a commitment to morality, there will be enough people to sustain moral practice: we will have adequate rational grounds to expect that our commitment to morality will be reciprocated. If en masse we were to stop caring about morality, this would indicate a radical change in our view about the kind of beings we are or the kind of social interaction we want. Moral reasons would cease to be reasons we have. But this does not support the inference that we would be obliged to see our caring about morality as rationally arbitrary. Instead, we could see that our caring about morality is contingent.

Moral criticism, directed to reasonable persons like us, does not thereby lose its point. We continue to care about morality and are prepared to avoid acting wrongly. This is not a matter of submitting to the dogmatic dictates of authority figures, tradition, or local custom but, rather, of recognizing the reasons there are to act morally. It is true that we may not be in agreement and may remain in disagreement, after inquiry and discussion, about what morality requires. There is no guarantee that we eventually will converge on a particular account of morality's content and authority. But we can agree in general that we have reason to do what morality requires. Admittedly, Dworkin, Smith, and others who worry that rationalism had better be true might be impatient with this response. A threat to morality seems to come not from persons already disposed to care about morality, who seek to clarify why morality should motivate them—the threat seems to come from intractable amoralists What can we say to amoralists if we have given up the position that moral requirements are categorical requirements of reason?

---

[18] Smith, "Dispositional Theories of Value," *Proceedings of the Aristotelian Society*, Supplement Volume LXIII (1989): 89–111, at p. 103.

My rejoinder is that we can only explain to amoralists why there is reason to act morally We realize that moral criticism will have little chance of convincing them that they should take morality seriously enough to be morally motivated This is not an embarrassment to morality as a normative endeavor Moral discourse is appropriate, if not efficacious, even when directed to rational beings who do not share our fundamental values Amoralists could still be capable of comprehending our commitment to morality sharing values is not always prerequisite to comprehending those values Furthermore, amoralists could recognize that there are reasons to act morally, without their having such reasons Their attitude toward morality might be similar to my attitude toward origami They might understand and respect why moral reasons motivate other people, while not being motivated by these reasons vis-à-vis their personal interests In short, amoralists might not be thoroughly beyond the reach of moral reasoning

Moral criticism can have various purposes Typically, it is directed to persons who have acted wrongly, or who have expressed beliefs or attitudes that are morally objectionable The purpose is to prompt the person criticized to reexamine her values or her commitment to them, or to recognize that the values she already accepts properly apply in the context in question Another purpose is to reaffirm the moral community's commitment to values that have been violated In this case, moral criticism might be directed less to the wrongdoer than to the moral community overall, of which the wrongdoer might or might not be member We want publicly to condemn racial hate crimes, for example, in order to remove any doubt within our society that race hatred is wrong, we might be quite pessimistic that the offender will accept this value Moral criticism also can have an expressive purpose, giving voice to our felt responses to wrongdoing Perhaps this offers the best understanding of reactive attitudes such as resentment

Punishment appears to raise a more serious problem The imposition of loss, pain, or confinement goes beyond condemnatory words or plain social ostracism This helps to explain the attraction of the idea that punishment entails the concept of desert Severe, freedom-infringing treatment of persons when they cannot credibly be said to deserve it is widely viewed as unjust If wrongdoers do not themselves have reason to avoid acting wrongly, we might wonder whether they can "deserve" punishment at all How, then, could amoralists—unlike the rest of us, who do have reason to avoid acting wrongly—be eligible for punishment on the grounds of desert? Perversely, their ultimate indifference to morality might seem to place them not only outside the moral community but also outside the reach of moral judgment that would legitimate severe treatment of them for violating certain moral norms

This problem could be addressed by showing that the permissibility of punishment need not rely on desert[19] If such an approach proves too controversial, we could give up conceptualizing freedom-infringing treatment of wrongdoers as "punishment" Instead, we could say that the practice has the aim of incapacitating or deterring persons who do unwarranted harm to others Such an aim seems reasonable, and it can sidestep the metaphysics of free will and moral responsibility that animate the issue of desert While we would prefer that our reasons in defense of the practice resonate with amoralists, its legitimacy does not depend on the morally salient reasons we have being reasons for them We can admit that the only reasons intractable amoralists would recognize are the prudential reasons they have to avoid being subjected to incapacitating or deterring treatment The point is that unwarranted harmdoing is generally unacceptable, no matter who does it, given the stakes for individual victims and society

Thus, rationalism does not need to be true in order for moral criticism or severe, freedom-infringing measures ever to be appropriate That all rational agents might not be bound by morality's claims does not mean that our commitment to the values underlying those claims is rationally arbitrary This explains why the justifiability of our caring about morality is not hostage to the fortune of debate in theoretical ethics about whether there are "moral facts" in a rationalist sense[20] At the same time, we should care whether our moral values have rational or reasonable credentials we want our ostensible moral reasons to be bona fide reasons That we feel deeply and strongly about the reasons we take to be moral reasons is not a bulwark against the possibility of dogmatism We can do better than this by working to develop a fuller account of the values that moral reasons reflect and of the relation between moral reasons and nonmoral values

LIONEL K MCPHERSON

Tufts University

---

[19] For critical discussion of desert with regard to punishment and moral responsibility, see Erin Kelly, "Doing without Desert," *Pacific Philosophical Quarterly*, LXXXIII (2002) 180–205 Like Kelly, I do not take skepticism about desert to rely on or substantiate compatibilism or, more specifically, a Strawsonian account of reactive attitudes

[20] Smith, *The Moral Problem*, p 202

PRESUPPOSITIONS AND SCOPE*

Within the philosophy of language, names and descriptions are traditionally taken to represent different paradigms of reference. Many philosophers draw a sharp contrast between the way speakers use names to talk about individuals and the way they use definite descriptions to do so. A proper name is used to pick out one specific individual. A definite description, on the other hand, provides a general formula for picking out distinct individuals in different situations. Metaphorically, a name is a tag attached to an individual, whereas a definite description is a set of instructions for finding an individual that satisfies some criterion.

This difference between names and descriptions is said to account for a well-known fact: descriptions exhibit narrow-scope readings with respect to modal operators while names do not. Here is an example in which a definite description has what is normally considered a scope ambiguity with a modal operator:

(1) Mary-Sue could have been married to the president.

Imagine (1) being uttered in a situation in which Grover Cleveland is the president. On one reading, (1) could be made true by a possible situation in which (a) Grover Cleveland is married to Mary-Sue and (b) Grover Cleveland is not president. This is the *wide-scope* reading of 'the president' since it picks out the individual satisfying the role in the actual world, regardless of whether he satisfies it in the possibilities considered. On another reading, (1) could be true because of a possible situation in which Mary-Sue is married to someone else, say Jake, who is president in that possible situation. This is the *narrow-scope* reading of 'the president' since the description picks up its referent within the possible situation considered.

Consider, by contrast, what happens if we replace the description in (1) with a proper name:

(2) Mary-Sue could have been married to Grover Cleveland.

---

* I am indebted to Eliza Block, Jessica Boyd, Sam Cumming, Delia Graff Fara, Gilbert Harman, Irene Heim, Nathan Klinedinst, Margaret Miller, Jim Pryor, Philippe Schlenker, Brett Sherman, and Edwin Williams for helpful comments on earlier drafts and/or discussion of these topics. I am also grateful to an audience at the 2005 Sinn und Bedeutung conference at Humboldt University in Berlin for many useful suggestions. An earlier version of some parts of this paper appeared in the proceedings of that conference edited by Christian Ebert and Cornelia Endress (Berlin: ZAS Papers in Linguistics, 2006).

0022-362X/07/0402/71–106

There is no way of understanding (2) as having two different readings analogous to those of (1) Even if, as a matter of their syntax, proper names can have different scope with respect to modal operators, there are no different truth-conditional readings corresponding to the different scopes the name can take

The standard picture of names and descriptions explains this difference between them Modal operators are generally taken to quantify over different possible situations Since names are tags linked to individuals while descriptions are instructions for finding an individual in a given situation, only the latter can pick out different individuals across different possible situations This line of reasoning forms the basis of Saul Kripke's famous modal argument for the claim that names cannot be semantically equivalent to descriptions [1]

This paper centers on a simple observation scope ambiguities between definite descriptions and modal operators are only sometimes available (or, at least, are only sometimes apparent) It turns out that the narrow-scope readings of definite descriptions within modal operators are only available when the common ground—the mutual beliefs of the conversational participants—includes the proposition that across a wide range of possible situations the descriptive content has a unique satisfier

One task of this paper is to give a principled explanation of this observation What sort of explanation is available depends on our understanding of the semantics and pragmatics of definite descriptions For this reason, it will be necessary to look at different accounts of definite descriptions I consider two basic views of descriptions the Russellian view, which treats definite descriptions as a kind of quantifier, and the presuppositional account, which treats them as expressions that presuppose the uniqueness or salience of an object (While these views can, in some respects, be reconciled, they also represent very different perspectives on descriptions) I argue that the Russellian view of descriptions, unaugmented, lacks the resources to predict correctly the availability of the narrow-scope reading of descriptions By contrast, presuppositional accounts of definite descriptions can make the right predictions To show this I discuss how presuppositions interact with modal operators

---

[1] Kripke, "Naming and Necessity," in Gilbert Harman and David Davidson, eds, *Semantics of Natural Language* (Boston Reidel, 1972), pp 253–355 The modal argument is widely discussed in the philosophy of language for example, Leonard Linksy, *Oblique Contexts* (Chicago University Press, 1983), Scott Soames, *Beyond Rigidity* (New York Oxford, 1986), and Jason Stanley, "Names and Rigid Designation," in Bob Hale and Crispin Wright, eds , *A Companion to the Philosophy of Language* (Malden, MA Blackwell, 1997), pp 555–85

The remainder of the paper relates this account of definite descriptions and modals to the contemporary debate about the semantics of proper names I argue that, once we adopt a presuppositional account of descriptions, the modal argument against descriptivist theories of names loses its force To make this argument, I consider a treatment of proper names which construes them as linguistic devices akin to definite descriptions According to this picture, both types of expressions are used to pick out individuals that satisfy some descriptive content I show that this account accurately predicts the behavior of names with respect to modal operators

The structure of this paper is as follows Section I presents the empirical data on the scope ambiguities of definite descriptions and other operators Sections II and III examine how the Russellian and presuppositional views of descriptions handle this data Section IV relates these facts about the scope of descriptions to the philosophical discussion about which propositions are expressed using sentences that include descriptions Sections V through VII consider the relationship between proper names and descriptions

I DESCRIPTIONS UNDER MODAL OPERATORS

First, we need to look at the details of the interaction of definite descriptions with modal operators The key observation here is that definite descriptions have distinct wide- and narrow-scope readings with respect to modal operators Although this observation plays a central role in much of the philosophical discussion of names and descriptions, there is little in the way of detailed study of the phenomenon [2]

It will be useful to think of modal operators—like 'must' and 'might'—as quantifiers over possible worlds (or situations) To say that something must happen is to say that in all possible worlds it does happen To say that something can happen is to say that there is a possible world (or situation) in which it does happen Of course, modality comes in different flavors modal operators may be read metaphysically, epistemically, or deontically In this paper, I will con-

---

[2] Within the semantics literature, most discussion of the interaction of descriptions and modals centers around the phenomenon of modal subordination Here is an example of modal subordination

A bear might come in to the cabin The bear would eat you

The modal in the second sentence, although universal in force, is only interpreted relative to the worlds involving the possibility mentioned in the first sentence, see, for instance, Craige Roberts, "Modal Subordination and Pronomial Anaphora in Discourse," *Linguistics and Philosophy*, XII (1989) 685–721 In this paper, I will not discuss either this phenomenon or *anaphoric* uses of definite descriptions like the use of 'the bear' in the second sentence, which refers back to the indefinite 'a bear' in the first sentence centrate on metaphysical modals—in keeping with much of the philosophical literature on names, descriptions, and modals.

Let us consider an example in order to get a grip on the narrow-scope readings of definite descriptions with respect to modal operators.

(3) Aristotle might not have been the teacher of Alexander.

If we read the modal as having a metaphysical force, it is natural to think that (3) is true. But since Aristotle was the teacher of Alexander, the sentence can only be true if the description 'the teacher of Alexander' picks up its reference under the modal operator. In other words, 'the teacher of Alexander' must pick out different individuals in the different worlds over which the modal operator quantifies. The truth of (3) is then established by the existence of a possible world in which the description 'the teacher of Alexander' picks out someone besides Aristotle. In that possible world, Aristotle is not the teacher of Alexander. By contrast, the wide-scope reading of the description could not possibly be true. This is because, on the wide-scope reading, 'the teacher of Alexander' picks out its referent in the actual world. But, in this case, it picks out Aristotle and the sentence would then assert that in some possible world Aristotle is not Aristotle, which is false.

Before moving on let me make a cautionary note. Sometimes the narrow-scope reading of a sentence containing a description and a modal may not be distinguishable from the wide-scope reading. If the sentence only quantifies over possible worlds across which one and the same person satisfies the description, it will be impossible to tell from the truth conditions of the sentence whether the description within it takes narrow or wide scope. For this reason, all of my claims about when we can or cannot get a narrow-scope reading of a sentence apply only to contexts in which the different scopes have an effect on the truth conditions of the sentence.

1.1. Role-type versus Particularized Descriptions. Example (3) in the previous section demonstrates that some definite descriptions have narrow-scope readings under modal operators. But the modal argument, as we shall see, relies on the claim that this is generally true of definite descriptions and this is the claim I wish to dispute. In order to do so, I need to make a distinction between two kinds of definite descriptions, which I call *role-type* and *particularized* descriptions.

One conceptual tool necessary to formulate this distinction is the idea of the *common ground*.[3] The common ground consists of all the propositions that are taken for granted by all the participants in a conversation. When I am talking to someone in a train station, for instance, the proposition that we are in a train station is generally part of the common ground. A proposition need not actually be believed to count as part of the common ground, but the participants must at least pretend to believe it or must use it as a working assumption. There is also an iterative aspect to the common ground. When a proposition is in the common ground it is not only mutually assumed to be true by participants of a conversation, but also mutually assumed to be mutually assumed to be true, and mutually assumed to be mutually assumed to be mutually assumed to be true, and so on.

---

[3] Robert Stalnaker, "Pragmatic Presupposition" in Milton K. Munitz and Peter K. Unger, eds., *Semantics and Philosophy* (New York: University Press, 1974), pp. 197–213; "Assertion" in Peter Cole, ed., *Syntax and Semantics*, Volume 9 (New York: Academic

A description is a *role-type description* if it is part of the common ground that there is exactly one person (or one salient person) satisfying the descriptive content across a range of relevant metaphysically possible situations and that the satisfier sometimes varies from situation to situation.[4] Some examples of role-type descriptions are 'the family lawyer', 'the mayor', 'the president', 'the tallest pilot', and 'the director'. With role-type descriptions, we usually know independently of the specific conversational situation that the descriptive content is satisfied uniquely across other possible situations. It is part of general knowledge that cities generally have one mayor, countries one president, and so on. Of course, many role-type descriptions are incomplete in the sense that they need to be augmented by an implicit specification of the particular role in question—so, for instance, 'the president' might be used to mean 'the president of the U.S.' or 'the president of the board of trustees'. Likewise superlative descriptions, such as 'the tallest man', require some domain within which they operate: 'the tallest man' might mean 'the tallest man in the room', or 'the tallest man in the galaxy'. But the basic criterion stands: a role-type description is a description for which it is part of the common ground both that the content of the (completed) description is uniquely satisfied across a wide range of possible situations and that the satisfier varies amongst these situations.

Particularized descriptions are simply those descriptions that are not role-type descriptions. The mark of a particularized description, then,

---

Press, 1978), pp. 315–32; "Common Ground," *Linguistics and Philosophy*, xxv (2002): 701–21.

[4] Note that while the number of metaphysically possible situations may be great, only certain situations are relevant when we use modals in normal speech with their metaphysical force.

is that it is not part of the common ground that the descriptive content has a unique but varying satisfier across a whole range of relevant metaphysically possible situations. Descriptions whose only content consists in general properties shared by many different individuals tend to be particularized descriptions, such as, 'the tall boy', 'the dog', and 'the loose-fitting cap'. Descriptions that refer to people by their physical location or what they did at some point are also usually particularized, such as, 'the man I met yesterday', 'the person over there', and 'the cat in the basement'. The reason these descriptions count as particularized—in ordinary contexts—is that we can only know that there is a single most salient individual satisfying the descriptive content (and thus the description picks some individual out) by having some sort of knowledge particular to the narrow conversational context (for example, for 'the tall boy' we must know that there happens to be exactly one tall boy around). I might further note that particularized descriptions may also be "incomplete" in the sense that one might naturally fill out descriptions like 'the tall man' with extra information such as 'in this room'.[5]

Whether a description counts as particularized or role-type depends upon what the common ground is. This means that corresponding to almost any particularized description there is some conceivable conversational context in which that description would count as a role-type description, and vice versa. So the distinction is not one between different types of linguistic expressions, but between different types of expression/context pairs. However, certain descriptions cast themselves more naturally as one sort or the other. When I give an example it will be clear if I mean it to be particularized or not.

It is worth noting that the role-type/particularized distinction is not the famous distinction between referential and attributive uses of descriptions introduced by Keith Donnellan.[6] On Donnellan's scheme, roughly speaking, attributive descriptions are used to speak of whoever satisfies the predicative content of a description, whereas referential descriptions are used to refer to known individuals. Whether a definite description falls on one side or the other of Donnellan's distinction depends on how it is used, how it is classified according to my distinction depends, instead, upon the relationship between the common ground and the predicative content of a description. Classification according to my distinction is independent of how a description is used, and, so, is independent of how it sits with regard to Donnellan's distinction. (But there may be points of contact. For instance, when a description is used attributively the conversational participants typically assume, or pretend to, that across different epistemically or different metaphysically possible situations different individuals would satisfy the descriptive content.[7])

---

[5] How incomplete descriptions are dealt with is a matter of much controversy within formal semantics and philosophy of language. The problem is discussed, among other places, in Scott Soames, "Incomplete Definite Descriptions," *Notre Dame Journal of Formal Logic*, xxvii (1989) 349–75.

[6] Donnellan, "Reference and Definite Descriptions," *Philosophical Review*, LXXV (1966) 281–304.

As it happens, there is interesting evidence of the linguistic significance of my distinction. Christopher Lyons reports that a northern Frisian dialect, Fering, seems to mark the distinction by using different articles to signify descriptions of the two sorts.[8] Here is what he says about the use of the two definite determiners—the D- and the A-articles—in Fering:

> In general the D-article is used where the identity of the referent is to be found by searching the spatio-temporal or textual context. The referent, or a prior reference to it, is there to be picked out, and in some cases distinguished from other entities satisfying the same description.... The A-article is used as a role-type one, and where, given the hearer's general knowledge or knowledge of the wider situation and of appropriate associations, the description is enough to single out the referent without the need for ostension (*ibid.*, p. 163).

Whatever one thinks of the D-article, the A-article is used to mark descriptions that seem a lot like role-type descriptions. So there is evidence that role-type descriptions form a natural pragmatic category capable of being marked in a language with its own article.

1.2. Role-type and Particularized Descriptions with Modals. Now, as we have seen, role-type descriptions allow narrow-scope readings with respect to modal operators, as in (3), repeated here:

(3) Aristotle might not have been the teacher of Alexander.

The description from (3), 'the teacher of Alexander' can easily be a role-type description since it can be part of the common ground that across a wide range of possible worlds Alexander would have had a teacher, but not necessarily the same teacher (for example, a different student of Plato might have been chosen instead to be Alexander's teacher). The question I turn to now is whether particularized descriptions exhibit the same sort of behavior with regard to modal operators as role-type descriptions do.

---

[7] I think one can generalize the notion of role-type and particularized descriptions to epistemically possible situations in addition to metaphysically possible ones, though I do not explore that here. Thus it may be that attributive uses are only possible with role-type descriptions. I discuss these points in greater detail in section IV.

[8] Lyons, *Definiteness* (New York: Cambridge, 1999).

Let us look at an example. Suppose that I went to a reception last night. At the reception, we can suppose, I talked to many different people for brief periods of time. Now, suppose that I learn that my old friend Hans was due to come to the reception but that he did not make it because his plane was delayed. Let us suppose that for this reason it is a relevant possibility that Hans could have made it to the reception, and that, if this were the case, I would have talked to him all night at the reception. This possible situation, if it were actual, is one which I could aptly describe with this sentence:

(4) Hans is the person I talked to the whole time.

Now suppose that I want to express to someone at the party that I consider (4) to be a possibility. One might think that I could do this by uttering a version of (4) with a possibility modal:

(5) Hans might have been the person I talked to the whole time.

There is, however, something very odd about using (5) to express the possibility of a situation in which (4) is true (assuming there is actually no one who I talked to the whole time). Indeed, if I utter (5) at the party, I will probably confuse my audience. (I will discuss a bit later how one might try to make sense of such utterances.) This oddness is quite surprising, however. If the definite description 'the person I talked to the whole time' can have scope within the modal operator, then we would expect that (5) would express the possibility of a situation within which (4) is true. Since such a situation is possible we would expect the utterance to be not only felicitous but also true. However, for some reason this narrow-scope reading of the description 'the person I talked to the whole time' is not actually available.[9] (The wide-scope reading of the description is quite hard to get as well since there is no person in the actual situation the description could refer to.)

Let us consider another example. Suppose that throughout an entire dinner party Siegfried does not eat anything, and is unique in this regard. Suppose that I have another friend, say Siegmund, who also would not have eaten anything if he had been at the dinner. Now, suppose I say something like this:

(6) I might have enjoyed talking to the person fasting through the dinner.

---

[9] Those familiar with presuppositions may not be surprised by this, since this is, roughly speaking, predicted by the presuppositional theory of descriptions.

It does not seem like I could mean anything but that I might have enjoyed talking to Siegfried by an utterance of (6) This is true even if it is possible that Siegmund could have come and Siegfried not come In this possible situation, of course, Siegmund would have been the only person fasting Nonetheless, it does not seem like (6) can easily express the proposition that there is a possible situation in which I would have enjoyed talking to whoever was unique in fasting at the dinner, Siegmund, Siegfried, or someone else entirely In this respect we cannot easily get the narrow-scope reading of the description 'the person fasting through the dinner'

We can, however, create conversational backgrounds within which 'the person I talked to the whole time' has a narrow-scope reading in (5) and 'the person fasting through the dinner' has a narrow-scope reading in (6) First, take (5) again

(5) Hans might have been the person I talked to the whole time

Suppose that it is part of the common ground that I generally talk to one person throughout an entire evening (because, for instance, I always start an argument with someone about politics which lasts the whole evening) In this case, I could utter (5) in order to express the proposition that if Hans had come he would have filled the role of being the person I talked to all night However, this is a case in which 'the person I talked to the whole time', which would usually be a particularized description, acts as a role-type description since it indicates a role which is uniquely filled across many relevant counterfactual situations[10]

The situation is similar for (6)

(6) I might have enjoyed talking to the person fasting through the dinner

If we can take it for granted that there is usually exactly one person fasting at such dinners, or that the organizers had intended to invite

---

[10] I can only think of one other circumstance in which the description 'the man I talked to the whole time' could have a nonrigid, narrow-scope reading in an utterance of (5) This other case is the one in which the description 'the person I talked to the whole time' has already been introduced in either its definite or indefinite form in the conversation For instance, instead of just saying (5), I might have said (A)

(A) I could have talked to a person the whole time Hans might have been the person I talked to the whole time

If I utter (A) it seems that the description in the second sentence can have a narrow-scope reading, and thus the utterance might express something true However, in this case, the definite description is anaphorically linked to the indefinite description that precedes it I want to put aside these anaphoric uses of descriptions as they involve the description inheriting properties from the original use exactly one person who would not eat, though not any specific person, then the narrow-scope reading of (6) is available However, without such an assumption the reading is very hard to get

We have seen, then, that in order to get a narrow-scope reading of a definite description we need to treat it as a role-type one Sometimes in response to an utterance the audience changes their assumptions, and, hence the common ground through the process of *accommodation*[11] This process of accommodation can lead the audience to treat a description as a role-type one even if prior to the utterance it is not part of the common ground that the description designates a role Here is an example in which such accommodation might occur Suppose I utter (7) when discussing a party I have just been to

(7) If I had gotten there earlier I might have been the person in charge of hats

My audience would not take me just to be asserting that if I had gotten to the party earlier I would, by myself, have taken charge of the hats Rather, they must also assume that across a whole range of different possible ways in which the party could have transpired there would have been one person who saw to the hats Making this assumption, through accommodation, the audience can then understand my assertion in (7) to be the assertion that if I had gotten to the party earlier I would have played the role of dealing with the hats

To understand better the behavior of descriptions within modal operators it is worth comparing sentences with particularized descriptions with sentences containing a typical role-type description Here is one

(8) Adlai Stevenson might have been the president

There is a natural reading of (8) on which the role-type description 'the president' has narrow scope It is true, for instance, if there are relevant possible worlds where Stevenson beats Eisenhower These are worlds in which Stevenson is "the president " But that sort of reading, that is, the narrow-scope one, is exactly the reading we do not find for (5), (6), or (7) without choosing backgrounds in which the descriptions act as role-type ones

1 3 Role-type/Particularized and Adverbs of Quantification Particularized and role-type definite descriptions also differ in their behavior with respect to adverbs of quantification, such as 'sometimes',

---

[11] See David K. Lewis, "Scorekeeping in a Language Game," *Journal of Philosophical Logic*, VIII (1979) 339–59, and Stalnaker, "Common Ground "

'usually', and 'generally' Role-type descriptions easily allow quantification over different satisfiers whereas particularized descriptions do not

(9) The best student is usually reclusive
(10) The mayor is usually a Republican
(11) The man in the car is usually nice
(12) The tall solider from France is usually violent

Both (9) and (10) exhibit two different readings corresponding to whether or not the descriptions get their reference under the adverb of quantification One reading of (9) has it that a particular student, who is the best, usually acts reclusively, on the other reading, most of the students who are the best, say in different years or classes, are reclusive (10) exhibits a similar ambiguity The situation is different with (11) and (12), which prefer to be read as being about one particular person For example, in (11), 'the man in the car' must pick out the actual person satisfying the description, unless it can be taken for granted that there is a entire class of relevant situations in which there is exactly one man in a car In other words, the description 'the man in the car' must act as a role-type one in order to get narrow scope Likewise, to get (12) to quantify over different individuals one has to imagine there being a recognized role the tall solider from France

The generalization is that role-type descriptions but not particularized descriptions can have narrow scope with respect to modal operators and adverbs of quantification The explanation of this observation requires us to delve into some of the details of different accounts of descriptions In the next section, I sketch the Russellian account and argue that it does not provide us with the tools needed to explain the data

II RUSSELLIAN DESCRIPTIONS AND SCOPE AMBIGUITIES

Russell's approach was to treat definite descriptions as a form of quantifier [12] While the indefinite description, 'an F' simply introduces existential quantification, $\exists xFx$, into a rendering of the meaning of a sentence in first-order logic, the definite description is more complicated [13] Russell claimed that definite descriptions are used to assert not just that there is an x such that Fx but also that there is only one x such that Fx. So, we can say that 'the F' expresses something equivalent to $\exists x(Fx \& \forall y(Fy \rightarrow y=x))$. I will abbreviate all this as $[\iota x: Fx]$. These quantifiers are introduced into a *logical form* of a sentence, which, for our purposes here, is a representation in logical notation (or something approximating logical notation) of a proposition that can be expressed by the sentence.

---

[12] A comprehensive discussion and defense of the Russellian view is given by Stephen Neale, *Descriptions* (Cambridge MIT, 1990)

[13] This treatment ignores some complications of Russell's actual position Russell did not really think 'a F' (or 'the F') was equivalent to any one part of the logical form of a sentence However, I am more interested here in the quantificational view inspired by Russell than in Russell's view itself

An important aspect of Russell's view of descriptions is that on this view definite descriptions can have scope at different places in a sentence. Russell's view, in fact, predicts certain systematic ambiguities in sentences containing a description and another quantifier. For, a critical component of the Russellian view is that a definite description introduces into the logical form of a sentence both a quantificational phrase, $[\iota x: Fx]$, and a variable x which is bound by it. This quantificational phrase can move within the logical form of a sentence, on the Russellian view, and this means that its scopal position with regard to any other quantifiers and operators in the sentence may vary. Such a sentence may be read in different ways depending on the scope that the description is taken to have.[14]

As an example consider this sentence:

(13) A man met the president.

The Russellian may render this sentence thus:

(14) $\exists x\, [\iota y: \text{president}\ y]\ x \text{ met } y$

Or alternatively, he may render it thus:

(15) $[\iota y: \text{president}\ y]\ \exists x\ x \text{ met } y$

The only relevant quantifiers at work here are existential quantifiers. Since commuting existential quantifiers has no effect on meaning, these different scopes for the definite description do not, in this case, correspond to different truth conditions for the sentence. But this is not always the case. As we shall see, the different scopal readings predicted by the Russellian often do correspond to different truth conditions for a sentence.

Now consider modal operators. We can treat modal operators as functions that take propositions and return propositions. There are two kinds of modal operators, the necessity operator, $\Box$, and the possibility operator, $\Diamond$ If we take these operators to have their standard meanings, $\Box$ returns a proposition that is true in a world *w* just in the case the input proposition is true in all the worlds possible with respect to *w* and $\Diamond$ returns a proposition that is true in a world *w* just in case the input proposition is true in at least one world possible with respect to *w*

---

[14] Recently many have attempted to integrate the Russellian account of quantifier scope with linguistic research into syntactic movement, along the lines of Robert May, *Logical Form: Its Structure and Derivation* (Cambridge: MIT, 1985). However, for our purposes these syntactically oriented reformulations of the Russellian view are not relevantly different from the standard Russellian view. This is especially true because there are no known syntactic restrictions on the scope of definite descriptions.

The Russellian view of descriptions in combination with this view of modal operators predicts that there will be certain systematic ambiguities in sentences that contain descriptions under modal operators Here is a sentence containing a description and a modal operator

(16) The president might have been wise

There are at least two different readings for (16) On one reading, it says of the actual president that there is a possible world in which he or she is wise This reading is captured by the formula (17) on which 'the president' takes wide scope over the modal operator

(17) $[\iota x\ \ \text{President}\ x]\ (\Diamond \text{wise}\ x)$

On the other reading of (16), the sentence is true just in case there is a possible world in which the person who is president within that possible world is wise, whoever he or she may be The logical form of this reading is captured by the formula in which the description gets scope under the modal operator

(18) $\Diamond\ [\iota x\ \ \text{President}\ x]\ (\text{wise}\ x)$

A major attraction of the Russellian view of definite descriptions is that it predicts the appearance of these readings by allowing quantifiers to have different scopes on different readings of a given sentence

But, it is this very mechanism of quantifier scope-taking that creates problems for the Russellian approach and its descendants The Russellian view makes the *prima-facie* prediction that, for a given sentence, both higher and lower scope with respect to a metaphysical modal operator will be available to a definite description We have seen already, however, that definite descriptions do not generally exhibit scope ambiguities with modal operators They do so only when the descriptions are role-type ones The Russellian view provides no obvious means for explaining why these different scopes are only sometimes available any mechanism generating these different scopes should, as far as the Russellian is concerned, apply equally well to particularized and role-type descriptions So the Russellian account, at first blush, makes the wrong prediction

Calcutta University Library ALIPORE CAMPUS

To make this argument clearer, it is worth considering a sentence with a description, which, in most contexts, lacks a scope ambiguity:

(19) I might have married the deserter.

Someone who accepts the Russellian account of descriptions should expect there to be a scope ambiguity between the two quantificational devices, 'the' and 'might'. The Russellian ought to anticipate both a wide-scope reading, on which the sentence says that I might have married the unique actual deserter, and a narrow-scope reading, on which it says that there is a possible situation in which I marry a person who is unique in deserting in that situation. But, this second reading is obviously not available, unless we build in enough background information to turn the description 'the deserter' into a role-type one.[15] So it turns out that (19) is ambiguous only in certain contexts (that is, only in combination with certain common grounds).

In response, the Russellian could suggest that the narrow-scope reading is unavailable in these contexts for pragmatic reasons; after all, not every reading of a conceivably ambiguous sentence is always found in every conversational context. But for an appeal to pragmatics to provide more than just a hand-wavy avoidance of criticism, a serious account of the data must be provided.

One such account might claim that the proposition expressed by the narrow-scope reading is only plausibly assertable in cases where the description is a role-type one. But I doubt that this is correct. Consider, for instance, these examples:

(20) Saint-Simon could have been the man I met at the zoo.
(21) Filibert could have been the person in court.

The Russellian predicts very sensible narrow-scope readings for both of these sentences. Let us restrict our attention to the metaphysical readings of the modals. The sentence (20), on the narrow-scope reading, asserts that Saint-Simon could have been a man I met at the zoo and been unique in being so; (21) says that Filibert could have been the only (relevant) person in court. Neither of these propositions is so peculiar as to be something one would not express. Nor do the situations in which one would express them require the descriptions to be role-type Yet, (20) and (21) do not exhibit narrow-scope readings of their definite descriptions, unless the descriptions act as role-type ones[16]

---

[15] Again, there are actually two ways of making a narrow-scope reading available. One is to create enough background information to make 'the deserter' a role-type description. The other is to precede (19) with a sentence like 'Someone might have deserted'. These two ways might amount to the same thing, but it is not obvious that they do. See footnote 2.

III PRESUPPOSITIONAL DESCRIPTIONS AND SCOPE AMBIGUITIES

Since the basic Russellian view does not give us any clear way of explaining the data discussed above we need to examine presuppositional accounts of definite descriptions In this section, I argue that a presuppositional account of definite descriptions explains the data concerning the scope of descriptions with respect to modal operators

3 1 Presuppositions and Descriptions I use *presupposition* to refer to an empirical phenomenon, not to any particular technical account of that phenomenon[17] When a proposition is presupposed by a statement, the statement seems to suggest or imply the proposition without exactly asserting it Take this sentence

(22) John went to the DMV too

An utterance of (22) seems to carry some sort of strong commitment to the proposition that someone besides John went to the DMV It would be ludicrous to assert (22) and to deny in the next breath that anyone besides John went to the DMV Yet (22) does not assert the proposition to which it is in this way committed—nor could it happily be used to assert this proposition Consider also this sentence

(23) My niece is a professional lutenist

As in the case of (22), an utterance of (23) would not seem, intuitively, to be an assertion that the speaker has a niece, though such an utterance would certainly imply it

---

[16] The wide-scope readings with metaphysical modals are also quite hard to get as they seem to assert that one person might have been another person Of course, the descriptions in (20) and (21) may get narrow-scope readings if we think of the modals in the sentences as having epistemic force I would argue that in these cases the descriptions act as role-types ones with respect to the epistemically possible worlds that the epistemic modals quantify over However, I cannot defend this view here

[17] Soames ("Presuppositions," in D Gabbay and F Guenther, eds , *Handbook of Philosophical Logic*, Volume IV (Boston Reidel, 1989), pp 553–616), and David Beaver (*Presupposition and Assertion in Dynamic Semantics* (Stanford CSLI, 2001)), review the linguistic evidence for presuppositions and the technical accounts of presupposition phenomena Some are skeptical that there is any unified phenomenon deserving the name of presupposition, but I think the large amount of excellent research on presuppositions such as that reviewed in the articles cited can put such worries to rest Nonetheless, there are undoubtedly differences between different kinds of presuppositions, such as those differences discussed in Mandy Simons, "On the Conversational Basis of Some Presuppositions," in R. Hastings, B Jackson, and Z Zvolensky, eds , *Proceedings of Semantics and Linguistic Theory 11* (Ithaca CLC Publications, 2001), pp 431–48

Expressions giving rise to this phenomenon—expressions like 'too' and 'my X'—are known as presupposition triggers. One striking feature of this class of expressions is that any presupposition that they trigger survives embedding under negation. Consider for instance the negation of (23):

(24) It is not the case that my niece is a professional lutenist,

Someone who utters (24) would still seem committed to the proposition that he has a niece. This phenomenon whereby a presupposition survives some operation on a sentence is called *presupposition projection*. And it occurs under other operators besides negation. Consider, for instance, the difference in semantic contribution between an assertion combined with a pronoun, 'I have a wife and she . . .', and the simple noun phrase, 'My wife . . .', in different contexts:

(25) a. If I had a wife and she were rich I wouldn't be in debt.
b. If my wife were rich I wouldn't be in debt.
(26) a. Do I have a wife and is she rich?
b. Is my wife rich?
(27) a. It's possible that I have a wife and she is rich.
b. It's possible that my wife is rich.

In (25b)–(27b) above, the presupposition that I have a wife survives the embedding within, respectively, a conditional, an interrogative, and a modal context. This accounts for the difference in meaning between all of the pairs above. It is the basic mark of presupposition triggers that the presuppositions they generate are projected in this way. But, the details of projection phenomena, and the exact characteristics of the contexts in which it does and does not occur, go far beyond the scope of this paper.

It is widely supposed that uses of definite descriptions trigger presuppositions. In fact, definite descriptions were the expressions first observed to exhibit presuppositional behavior and the expressions for which the notion of presupposition was developed.[18] On the standard view, definite descriptions trigger the presupposition that there is a unique satisfier of the descriptive content in the contextually relevant domain. On a competing view, definite descriptions trigger the presupposition that there is a maximally salient item satisfying the descriptive content in the contextually relevant domain.[19]

---

[18] Particularly in P. F. Strawson's, "On Referring," *Mind*, LIX (July 1950): 320–44.

[19] See Lewis, "Scorekeeping." Irene Heim (*The Semantics of Definite and Indefinite Noun Phrases*, Ph.D. Thesis, Massachusetts Institute of Technology, 1982) gives a similar view within the framework of formal semantics. Zoltán Szabó ("Descriptions and

The differences between the two accounts of the content of the presuppositions of definite description are not very important for this discussion, and I will often switch between them

I want to put aside, in this context, the case of felicitous uses of nonreferring definite descriptions such as this

(28) I didn't see the king of France at Versailles, because there is no king of France

There is much to say about such cases, but they do not challenge the basic observation that definite descriptions generally give rise to presuppositions of existence and uniqueness/salience [20] For one thing, we can say that in sentences like (28) the presupposition is explicitly contradicted, and that when this happens certain presuppositions disappear

There is a further question we might ask about the presuppositions of definite descriptions besides the question of their content Any account of definite descriptions must say something about what it is about definite descriptions that makes them trigger these presuppositions One might think that the fact that definite descriptions trigger presuppositions of existence and uniqueness simply follows from the Russellian treatment of descriptions discussed in the previous section However, the situation cannot be so simple For on the Russellian view, definite descriptions can be used to assert the existence and uniqueness of something satisfying the descriptive content A defender of the simple Russellian view needs to explain why the use of definite descriptions give rise to presuppositions with their particular projection behavior Some have suggested that the presuppositional behavior of descriptions can be predicted from their Russellian truth conditions by general pragmatic principles [21] However, as far as I know, no one has substantiated this claim and there are serious obstacles to it For instance, it is very hard for the Russellian to explain on pragmatic grounds the difference in communicative force between definite descriptions and explicit assertions of existence and uniqueness

---

Uniqueness," *Philosophical Studies*, CI (2000) 29–57), and Roberts ("Uniqueness in Definite Noun Phrases," *Linguistics and Philosophy*, XXVI (2003) 287–350), have recently advocated novel versions of this view

[20] Kai von Fintel ("Would You Believe It? The King of France is Back!" in Marga Reimer and Anne Bezuidenhout, eds , *Descriptions and Beyond An Interdisciplinary Collection of Essays on Definite and Indefinite Descriptions* (New York Oxford, 2004), pp 269–96) discusses this problem and refers to much of the major literature on it

[21] See Neale, *Descriptions*, p 54

when embedded in conditionals. Here is a minimal pair to illustrate this problem:

(29) If the king signed my pardon, I would be spared.
(30) If there were one and only one king and he signed my pardon, I would be spared.

Clearly (29) but not (30) is best uttered in a situation in which it can be assumed that there is a unique king. If descriptions, however, can be used to assert existence and uniqueness it is unclear why (29) but not (30) is inappropriate if it cannot be assumed that there is a king.

In response to such problems, we might instead treat the presupposition that a definite description triggers as an additional feature distinct from its normal truth-conditional content. In this case, we could maintain the Russellian account of definite descriptions and tack on to it the claim that definite descriptions also trigger presuppositions of existence and uniqueness. There are two problems with such a move. First of all, it leads to a redundancy in our account of definite descriptions since we must now take them to both presuppose and assert existence and uniqueness. Second, the account does nothing to explain why definite descriptions trigger the particular presuppositions that they do.

In response to these problems we might instead suppose that a definite description is a term whose function is to pick out the thing that satisfies the presupposition (that is, the unique/uniquely salient thing satisfying the descriptive content). In other words, we might suppose that [the F] is equivalent to [*x* s.t. *x* is the thing maximally salient in satisfying F].[22] I will call this view the presuppositional view of descriptions. On it, the presupposition of a description is not an extra component of the meaning of the description but rather arises from the fact that the semantic value of a description is undefined when the presupposition is not satisfied.[23] On this view, the function of a definite description is to refer to a maximally salient item; if it cannot do that, then its use is infelicitous. The task then left to the theorist is to explain the projection behavior of the presuppositions of descriptions based on this sort of story. There is much promising work in this direction, by Irene Heim,[24] Rob van der Sandt[25] and others. I will not, however, go into the details of presupposition projection except for the case of definite descriptions appearing under modal operators and adverbs of quantification.

---

[22] But I still wish to remain neutral on the question of whether descriptions are really object-referring expressions or rather contribute descriptive modes of presentation, an issue I turn to in the next section.

[23] There are both semantic and pragmatic versions of this view, as discussed in Stalnaker, "Pragmatic Presuppositions," but the differences between them are not important here.

A note. In my earlier discussion of the relationship between descriptions and modal operators, I treated the different readings of definite descriptions under modal operators as if they were the result of quantifier-scope ambiguities. One might wonder, however: If descriptions are not standard quantifiers, how can they nonetheless have different scopes with respect to other operators? This is basically a technical question and so its answer is also technical.[26] I will continue to refer to the different readings of sentences with descriptions and modal operators as the wide- and narrow-scope readings even if they are not manifestations of the usual quantifier scope ambiguities posited by the Russellian.

---

[24] Heim, "On the Projection Problem for Presuppositions," in Michael Barlow, Daniel Flickinger, and Michael Wescoat, eds., *WCCFL 2: Second Annual West Coast Conference on Formal Linguistics* (New York: Oxford, 1983), pp. 114–25.

[25] Van der Sandt, "Presupposition Projection as Anaphora Resolution," *Journal of Semantics*, ix (1992): 333–77.

[26] There are a variety of different means of representing the different scopes descriptions can take besides the usual Russellian way. Here are two methods. First, we can think of descriptions as including a world variable that can, but need not, be bound by modal operators or adverbs of quantification, as in Heim, "Artikel und Definitheit," in Arnim v. Stechow and Dieter Wunderlich, eds., *Semantics: An International Handbook of Contemporary Research* (New York: de Gruyter, 1991), pp. 487–535. An approximation of the logical form of a narrow-scope reading of sentence (A) would be represented by something like (B), where the description's semantic value is in brackets. (Again, I am not trying to take a stand on the issue of what exactly the semantic value of the description is—in particular whether it is exhausted by the object it refers to or rather involves a mode of presentation. I elaborate on these issues in section iv below.)

(A) The mayor could be a woman.

(B) $\exists$ world w, s.t. [the unique mayor in w] is a woman.

On the wide-scope reading we simply attach the description to the actual world, @, instead of letting it be bound by the modal operator:

(C) $\exists$ world w, s.t. [the unique mayor in @] is a woman.

Alternatively, we can use the approach of van der Sandt and Beaver and treat the presupposition itself as somehow scoping to different positions while binding a variable that is left in the place of the description. In the following examples, I use x to represent this variable, while putting the presupposition in parentheses and allowing its location to determine where it gets satisfied. Here is the narrow-scope reading, with the presupposition staying within the modal operator:

(D) $\Diamond$ ($\exists x$: x is a unique mayor) x is a woman.

On the wide-scope reading, the presupposition simply comes out of the modal operator into the front of the sentence:

(E) ($\exists x$: x is a unique mayor) and $\Diamond$ x is a woman.

3.2. Modals, Adverbs of Quantification, and Presuppositions. My goal in this section is to explain the fact that only role-type descriptions can have narrow scope under modal operators. I will do this by appealing to the presuppositional view of definite descriptions I have just presented.

First, I want to illustrate the general idea by looking at adverbs of quantification—expressions like 'always' and 'usually'—and then I will move on to modal operators. Adverbs of quantification can be treated as quantificational expressions that are very similar to modals. While modal operators quantify over possible worlds, adverbs of quantification quantify over situations—where a situation is something smaller than an entire world.[27] A sentence of the form 'always *x*' is, on this approach, equivalent to a sentence of the form 'in all situations, *x*'. (Just as, on the usual quantificational approach to modality, 'it must be the case that *x*' is equivalent to 'in all possible worlds *x*'.)

As we saw in section I.3, definite descriptions exhibit scope ambiguities with respect to adverbs of quantification. Consider, for instance, this sentence:

(31) The mayor is always happy.

On the wide-scope reading, (31) is used to make a statement about the current mayor and says that in all (relevant) situations he or she is happy. On the narrow-scope reading, (31) is used to assert that in any relevant situation, whoever the mayor is in that situation, he or she is happy.

Now, as I explained above, any use of a definite description triggers some sort of presupposition, perhaps a presupposition to the effect that there is a uniquely salient satisfier of the descriptive content. Whether a particular use of a definite description is felicitous or not will depend on whether the presupposition is part of the common ground (that is, whether it is one of the mutual assumptions of the conversational participants) or can be added to the common ground through accommodation. In other words, what a speaker presupposes, his audience must be capable of assuming.

When a definite description appears under an adverb of quantification we add, so to speak, an extra layer of presupposition. To see how this is so, consider first the narrow-scope reading of the definite description 'the mayor' under the adverb of quantification 'always' in (31) According to the semantics of adverbs of quantification I am using here, to determine whether the sentence on this reading is true we would need to check the truth of the embedded sentence ('the mayor is happy') in all the relevant situations In other words, the truth of the narrow-scope reading of (31) depends on the truth of (32) over a class of relevant situations

---

[27] See, for instance, von Fintel, "A Minimal Theory of Adverbial Quantification," in Barbara Partee and Hans Kamp, eds., *Context Dependence in the Analysis of Linguistic Meaning: Proceedings of the Workshops in Prague* (Stuttgart: IMS Stuttgart Working Papers, 1995), pp. 153–93. I am making a departure from Lewis's ("Adverbs of Quantification," in Edward Keenan, ed., *Formal Semantics of Natural Language* (New York: Cambridge, 1975), pp. 3–15) classic account of adverbs of quantification.

(31) The mayor is always happy
(32) The mayor is happy

Now, normally, any use of 'the mayor' triggers the presupposition that in the relevant circumstance there is one salient satisfier of the descriptive content, that is, that there is one salient mayor However, the narrow-scope reading of (31) does not just include one use of the description 'the mayor' By this I mean that the description is not just used once to pick out an object Rather an utterance of (31), on the narrow-scope reading of the description, makes use of the description to pick out an object in every situation being quantified over So the presupposition of (31), on its narrow-scope reading, is that across all the relevant situations there is a unique mayor Thus, when one intends (31) to have this narrow-scope reading, the audience must assume, or be capable of assuming, that this presupposition is satisfied (that is, that across every relevant situation there is a unique mayor)[28]

The situation is different with the wide-scope reading of the description in (31) For if we intend (31) to have this reading we pick out the mayor salient in our current context and assert that he or she is happy in all the relevant situations In this case the description is only being put to one use in the sentence it picks out the actual mayor So this sentence does not generate any extra presuppositions through the interaction of the definite description and the adverb of quantification The sentence is fine as long as the common ground includes the proposition that there is a salient mayor, or as long as that proposition can be accommodated

The critical point here is that, generally, a sentence for which the definite description is read as having wide scope with respect to an adverb of quantification triggers a different presupposition from a sentence for which the definite descriptions is read as having narrow scope. On the narrow-scope reading, the sentence triggers the presupposition that across all the relevant situations there is a unique satisfier of the descriptive content. On the wide-scope reading, by contrast, the sentence just triggers the presupposition that in the actual situation there is a unique satisfier of the descriptive content.

---

[28] This point also follows from any view which treats definite descriptions as having partially defined semantic values If one treats definite descriptions as undefined when they have no unique referent (as on the semantic approach to presupposition), then (31) on the narrow-scope reading will be undefined unless there is a unique mayor in all the relevant situations

My central claim is that the different presuppositions of the sentence on the narrow- and wide-scope readings, respectively, determine which readings are available in a given conversational context. When a sentence is ambiguous and its different possible readings have different presuppositions, then we are only able to get the reading(s) whose presupposition can be accommodated. Now, a particularized description is a description for which it is not assumed that there is a unique satisfier across the relevant situations. If the relationship between the common ground and the descriptive content has this feature then one cannot, without accommodation, get the narrow-scope reading of the description under an adverb of quantification. In such a case either the common ground must be changed (making the description a role-type one) or the other, wide-scope reading will be the only one available.

Consider, for instance, this sentence.

(33) The man from Apple is often troublesome.

For 'the man from Apple' to have narrow scope under the adverb of quantification in (33), the common ground needs to entail that there is a class of relevant situations in each of which the description picks out a uniquely salient person. If, for instance, it is part of the common ground that we often have meetings where Apple sends exactly one person, then the narrow-scope reading is possible. But this circumstance is one in which 'the man from Apple' functions as a role-type description. When 'the man from Apple' acts as a particularized description, by contrast, only the wide-scope reading is possible, that reading just requires that there be a uniquely salient man from Apple available in the actual situation.

Does this approach also cover the data concerning the scope of definite descriptions under modal operators? I think that it does. Recall that modal operators quantify over possible worlds or possible situations. Consider this sentence.

(34) Susan could have been married to the man wearing a big red hat.

This sentence fails to have a narrow-scope reading in normal contexts when the modal has metaphysical force. (34) cannot easily be read to mean that Susan's actual husband might have been unique in wearing a big red hat.[29] The presupposition of the sentence with the narrow-scope reading is that in each relevant possible world there is a unique, but potentially varying satisfier, of the description.[30] However, in normal situations there is not an easily accessible set of possible worlds such that in each world some (different) person is wearing a big red hat. Of course, there are many metaphysically possible worlds where just one person is wearing a big red hat (in the relevant situation), but they do not form a natural modal domain that we can quantify over with modal operators. Moreover, the presupposition of the sentence on the narrow-scope reading, that is, that in every relevant possible world there is unique person wearing a big red hat, is a bit too peculiar to come easily into play through accommodation, explaining our preference for the wide-scope reading. And, if we nonetheless accommodate this presupposition, the description becomes a role-type one.

The situation is different when we have, under a modal operator, a description like ‘the president’. Here, the narrow-scope reading is easily accessible.

(35) The president of the company might have been nicer.

The narrow-scope reading is available because we can naturally think that over a relevant set of possibilities the company would have had some president or other in each possibility. When we think of the modal as quantifying over these possibilities, the description ‘the president of the company’ can get narrow scope with respect to the modal.

Let me review what I have been arguing. First, the Russellian account does not predict the systematic scope variations definite descriptions show with respect to modal operators. This is because the Russellian should, all else equal, expect that in any sentence with a modal operator and a description there is a scope ambiguity between them. Moreover, the limited distribution of readings that we find cannot easily be explained by pragmatics alone.

Second, presuppositional accounts of descriptions can explain the distribution we find. Adverbs of quantification and modal operators quantify over sets of situations (or worlds) Descriptions can get narrow scope if and only if the presupposition that there is a unique satisfier in every situation is satisfied If this presupposition is satisfied, or can be accommodated, then we can get the narrow-scope reading, otherwise, we only get the wide-scope reading on which the description refers to an object in the current context

---

[29] It could have a natural narrow-scope reading if we read the modal epistemically, however, epistemic modals raise different issues since the sets of worlds they quantify over are radically different.

[30] This assumes that the truth of the sentence depends upon what happens in all possible worlds, rather than just one possible world. This fact does not follow directly from the semantics of possibility modals since they are used to make an assertion of the form ‘there exists a possible world such that . . .’ However, I would argue that, for semantic or pragmatic reasons, possibility modals are nonetheless used to make statements about entire sets of possible worlds.

IV DEFINITE DESCRIPTIONS AND PROPOSITIONAL CONTENT

Before moving on, I want to relate what I have said about definite descriptions to some general issues within the philosophy of language Much of the work there is situated within a theoretical framework which derives from Frege and Russell One of the central features of this framework is the view that utterances of sentences in contexts are used to communicate abstract objects called *propositions* The philosophical literature on definite descriptions and proper names—the relation between the two being the subject of the remainder of this paper—focuses on the contribution these expressions make to the propositions expressed by utterances of sentences including them On one popular conception, sentences containing singular terms can sometimes be used to communicate object-dependent propositions These are propositions about some particular object To put it graphically, the object itself—rather than just properties of the object—is part of these propositions For instance, one might think that if there is a woman standing a few feet in front of me, I can say something like (36) to communicate a proposition about the woman herself, rather than about whoever might satisfy the description I use

(36) The woman in front of me may be a journalist

Other uses of singular terms in sentences do not communicate propositions about any particular object but speak about an object only by specifying some property it has For instance, one can talk about 'the owner of this car, whoever that might be' I will call propositions that are not about particular objects, *general* propositions

The question I want to take up here is how the preceding discussion of the potential scope of definite descriptions relates to the philosophical discussion of the contribution of definite descriptions to propositions expressed using them[31] Let me first recall that the distinction between particularized and role-type descriptions does not simply amount to the distinction between descriptions used to express object-dependent propositions (which are called *referential* uses of descriptions) and descriptions used to communicate general propositions (which are called *attributive* uses of descriptions) For instance, 'the president of the U S ' can be used to communicate an object-dependent proposition even when it functions as a role-type description

---

[31] Donnellan, "Reference and Definite Descriptions", Kripke, "Speaker's Reference and Semantic Reference," *Midwest Studies in Philosophy*, XXIX (1977) 255–76

(37) The president of the U S likes madeira

An utterance of (37) is potentially capable of expressing an object-dependent proposition about the current president of the U S So, it is not restricted to expressing a general proposition about the person happening to satisfy the description, despite the fact that the description itself is role-type [32]

Likewise, consider a situation in which (A) my potato salad was stolen in the middle of the night, (B) I am sure that one person stole it, (C) I am sure that if the thief had not stolen it no one else would have, and (D) I have no idea who the thief is

(38) The thief obviously liked Worcestershire sauce and thyme

In these circumstances, by uttering (38), I may communicate a proposition about whoever satisfies a certain property, that is, the property of being the unique thief in the current situation But, the description is a particularized one, since in different (relevant) possible situations either the same (unknown) person satisfies the description or no one does

Nonetheless, the difference between narrow- and wide-scope uses is correlated, to an extent, with the distinction between referential and attributive uses of definite descriptions Any description which in a given case receives narrow scope with regard to a modal operator is generally not used to express an object-dependent proposition This is because there is not one object such a sentence is used to talk about, rather the description seems to pick out different objects in different possible situations [33] Wide-scope uses of definite descriptions need not, on the other hand, express object-dependent propositions In the context of the potato salad calamity above I could also have said

(39) The thief might have left a little potato salad for our breakfast

---

[32] Whether such uses of descriptions to express object-dependent propositions about individuals follow from the semantics of descriptions or rather depend on various pragmatic mechanisms, as Kripke suggests, is an entirely separate issue

[33] Unless, of course, the narrow-scope reading picks out the same person in all the possible situations in which case it could be used to communicate a singular proposition

There is obviously a wide-scope reading of the description 'the thief' with respect to the modal 'might' However, this reading need not correspond to a communication of an object-dependent proposition about some individual Even with the wide-scope reading, (39) can be used to express a general proposition about whoever stole the potato salad

So merely establishing when different scopes are available is not enough to establish what the basic propositional contribution of descriptions are Indeed, I want my proposal to remain neutral on exactly this point Whatever account of the semantics and pragmatics of descriptions we give, it must explain both the referential and attributive uses of descriptions The current standard theory within philosophy takes descriptions to be semantically encoded to express general propositions and explains the referential uses by appeal to pragmatics[34] I have nothing to say about the merits of this proposal What I will argue for in the remainder of this paper is that, contra the philosophical orthodoxy on this question, facts about scope give us no reason to treat proper names any differently from definite descriptions

V PROPER NAMES AND THE MODAL ARGUMENT

An extremely influential argument in the philosophy of language, Kripke's modal argument, purports to show that proper names are not semantically equivalent to definite descriptions[35]

Here is one version of the argument

1 Definite descriptions exhibit narrow-scope readings with respect to modal operators
2 Proper names do not exhibit narrow-scope readings with respect to modal operators
3 The meaning of a proper name cannot be the same as that of a definite description

The argument depends upon the sort of observations I made in the introduction to this paper Consider, for instance, (2), repeated here

(2) Mary-Sue could have been married to Grover Cleveland

The modal argument begins by suggesting that on the hypothesis that 'Grover Cleveland' is really semantically equivalent to some description, 'the F', one should expect to find two possible readings of (2), corresponding to whether the description, 'the F', gets its scope under the modal (finding the satisfier of the description within each possible situation) or outside the modal (picking out its actual satisfier, that is, Grover Cleveland) However (2) does not seem to exhibit different readings of this sort So, the argument concludes, 'Grover Cleveland' cannot be equivalent to 'the F'

---

[34] Kripke, "Speaker's Reference and Semantic Reference", Neale, *Descriptions*

[35] Three pieces that seem to understand the argument this way are Linsky, *Oblique Contexts*, chapter 7, Stanley, "Names and Rigid Designation", Soames, *Beyond Rigidity*, chapter 2

Many have noted that the argument only shows that proper names are not semantically equivalent to those descriptions whose descriptive content allows them to pick out different objects in different possible situations In other words, the argument shows that proper names are not equivalent to those descriptions whose descriptive content is actually capable of being satisfied by different individuals in different situations Some descriptions do not have this property These include certain descriptions, such as 'the actual president', whose descriptive content contains some indexical reference to the actual world No matter what their scope is, such descriptions always pick out the same individual (they are so-called *rigidified descriptions*) In light of this qualification, we can view the modal argument as purporting to establish that, if proper names are semantically equivalent to any definite descriptions, they are semantically equivalent to rigidified descriptions like 'the actual mayor' [36]

The first premise in my presentation of the modal argument above states that definite descriptions exhibit narrow-scope readings with respect to modal operators In this paper, however, I have presented and explained a significant qualification to this claim I have shown that only *role-type* descriptions can have narrow scope with respect to a modal operator Thus, we need to revise our assessment of the modal argument in light of this qualification [37]

In fact, once we recognize that definite descriptions do not *always* exhibit scope ambiguities with respect to modal operators, the modal argument loses much of its force against descriptivist accounts of names. If one assumes that descriptions always exhibit scope ambiguities, then *one* instance of a sentence containing a proper name and a modal operator that does not show a scope ambiguity will serve to demonstrate that names cannot be descriptions (except perhaps rigidified descriptions). But, once we have recognized that descriptions do not generally show scope ambiguities, we can no longer reason in this way. Many definite descriptions, such as 'the man in the corner' and 'the person I saw yesterday', have restrictions on what scope they can get with respect to modal operators. These descriptions belong to the large class of descriptions that are particularized in most contexts and, thus, do not exhibit narrow-scope readings in these contexts. The modal argument fails to show that proper names are not equivalent to *these* sorts of descriptions.

---

[36] Discussion of rigidified descriptions include Stanley, "Names and Rigid Designation", Soames, *Beyond Rigidity*, and Michael Nelson, "Descriptivism Defended," *Noûs*, xxxvi (2002) 408–35

[37] Kripke explicitly acknowledges the degree to which his argument depends on descriptions acting Russellian, and hence being able to get narrow scope (*Naming and Necessity*, p 6, fn 8) Bart Geurts ("Good News about the Description Theory of Names," *Journal of Semantics*, xiv (1997) 319–48) also picks up on this issue, arguing that names are like certain descriptions which always take wide scope (though he does not offer an account of *why* these descriptions take wide scope)

> The presuppositions triggered by names seem to have a decidedly stronger tendency to 'take wide scope' than some others In this respect, too, they are on a par with other descriptively attenuate 'incomplete', definite descriptions like 'the door' or anaphoric pronouns like 'it' (pp 335–36)

It is worth noting that this is a significantly larger qualification than the one in the previous section about rigidified descriptions. Descriptions like 'the man in the corner' are not rigidified descriptions since they have a predicative content which different individuals can satisfy in different situations. So, the class of definite descriptions that are generally particularized includes descriptions which are not rigidified. In addition, while it is extremely hard to find real English expressions that act as rigidified descriptions ('the actual mayor' certainly does not), it is extremely easy to find English expressions that are usually particularized descriptions.

Another way of putting my basic point is to say that the modal argument still leaves open the possibility that names are particularized descriptions. Of course, whether a description is particularized or role-type depends upon the relationship between the common

---

Likewise, Dummett thinks that the modal argument lacks force because only some descriptions can have narrow scope within a modal operator. However, for some reason, he thinks the only descriptions that can take narrow scope are ones that appear after the copula:

> We are now in a position to understand the grain of truth in Kripke's doctrine of proper names as rigid designators. For modal contexts in general, there is no relevant difference between proper names and definite descriptions: but the matter stands otherwise when the name of the descriptions is preceded by the verb 'to be' or 'to become'.
>
> —Dummett, *Frege: Philosophy of Language* (Cambridge: Harvard, 1973), p. 131.

This is wrong, however, as the following minimal pair shows:

(A) John might have made friends with the president.
(B) John might have made friends with Grover Cleveland.

Only in the case of the first sentence can one detect a scope ambiguity. So definite descriptions do not only differ from proper names after the copula.

ground and the predicative content of the description. So, a name is unlikely always to be a particularized description, but a name might be equivalent to a definite description that has a descriptive content which makes it particularized in most contexts. This hypothesis would explain the resistance names show to taking narrow scope in most instances.

In the remainder of the paper, I examine one particular descriptivist conception of names to see whether, according to this conception, names can be construed as particularized descriptions. I will also look at contexts in which, according to this descriptivist proposal, names do not act as particularized descriptions. By looking at these contexts we can assess whether, as the descriptivist should predict, names can sometimes get narrow scope with respect to modal operators. I will argue that—contrary to the philosophical orthodoxy—the descriptivist view does extremely well at predicting the potential scope of proper names with respect to modal operators.

VI. NAMES AS METALINGUISTIC DESCRIPTIONS

The view that names are semantically equivalent to definite descriptions is often described as the view that names are *disguised descriptions*, since unlike real definite descriptions names do not openly show their descriptive content. This leads to the question of what the descriptive content of a name is. Here I will sketch one answer to this question, but I will not systematically consider alternatives.

One of the most plausible instantiations of the view that names are descriptions treats names as "metalinguistic" descriptions.[38] On this account, the meaning of a name *N* is roughly captured by the description 'the bearer of *N*'. We must distinguish this account of the semantics of proper names from the truism that a name *N* refers to whoever is referred to by *N*. The view that names are metalinguistic descriptions, unlike this truism, is neither trivial nor circular. We have a social practice of naming, under which one cannot bear a name just in virtue of some person using it to refer to you. So the facts about name-bearing are not mere trivial metalinguistic ones, like the fact that 'jump' means jump. In fact, the metalinguistic view of names makes a very strong claim: that each proper name has the same meaning as some particular definite description.

---

[38] Such views have a long tradition. Williams Kneale ("Modality De Dicto and De Re," in Ernest Nagel, Patrick Suppes, and Alfred Tarski, eds., *Logic, Methodology and Philosophy of Science: Proceedings of the 1960 International Congress* (Stanford University Press, 1962), pp. 622–33) explicitly advocates a metalinguistic view, and Burge ("Reference and Proper Names," this JOURNAL, LXX, 14 (August 16, 1973): 425–39) comes close to this view, though he treats names as predicates. See also Geurts, "Good News about the Description Theory of Names", Jerrold Katz, "The End of Millianism," this JOURNAL, XCVIII, 3 (March 2001): 137–66, and Kent Bach, "Giorgione Was So-Called Because of His Name," *Philosophical Perspectives*, XVI (2002): 73–103.

Definite descriptions trigger presuppositions, so, if proper names are equivalent to certain definite descriptions they will also trigger presuppositions. Earlier, I suggested that definite descriptions trigger the presupposition that there is a uniquely salient individual satisfying the descriptive content. So, if a name *N* were equivalent to the description 'the bearer of *N*', then a use of *N* would trigger the presupposition that there is a uniquely salient person bearing *N*. It seems plausible that names carry this presupposition. For when we use a proper name usually we presuppose that there is a most salient person bearing the name. Without this presupposition we could not expect our audience to understand to whom we meant to refer.

Kripke in *Naming and Necessity* makes other powerful arguments, besides the modal argument, against the view that names are disguised descriptions. His strongest argument, to my mind, is one about speaker knowledge. Here is a version of this argument. If the name 'Plato' were synonymous with the description 'the author of *The Republic*', then one would think that competent users of the name would have to know—at least implicitly—that Plato is the author of *The Republic*. However, it absurd to suppose that it is a condition on semantic competence with the term 'Plato' that one know that "Plato" wrote *The Republic*.

I do not think the metalinguistic view succumbs to this argument about speaker knowledge. The knowledge that a person referred to by a name bears that name may well be part of every competent speaker's grasp of the meaning of the name. The only objection to this that I can see is the claim that children are able to use proper names without having sufficient conceptual resources to grasp descriptions like 'the bearer of *N*'. There are a few things to be said about this. First, the conceptual capacities of very young children may be extremely sophisticated, so that the empirical claim may simply be false: children might, from their first uses of proper names, be in a position to grasp (in some sense) the descriptions associated with names.[39] Second, even if children can use proper names without grasping the descriptions associated with them, this does not mean that the adult use of proper names is not descriptive in the way I have suggested[40] Third, it may be that children's use of proper names is in some way parasitic on adult usage or deferential to it, so that if adults did not use names as metalinguistic descriptions children would not be able to use them to refer people at all These considerations show that the knowledge argument may not be successful against the metalinguistic view[41]

---

[39] Paul Bloom (*How Children Learn the Meaning of Words* (Cambridge: MIT, 2001)) discusses what conceptual capacities children might need to learn the meaning of names and other words.

VII NAMES AS DESCRIPTIONS UNDER MODALS

Now that we have a reasonable account of the descriptive content of proper names in hand we can see whether it predicts that names are usually particularized descriptions Recall that particularized descriptions are ones whose descriptive content is not commonly known to be uniquely satisfied by different individuals across a range of relevant possible circumstances It seems to me that in most contexts metalinguistic descriptions must be particularized For instance, it would require a very odd context to make it plausible that over an entire range of different possibilities there would always be a uniquely salient 'Samuel' available, but without this being the same person in each situation In many possible situations there is at least one person called 'Samuel', but it is hard to see why there would always be one most salient such person

In other words, metalinguistic descriptions are particularized definite descriptions in most contexts, since for most relevant classes of possible situations one cannot suppose there will be a different uniquely salient person satisfying the descriptive content in each situation Supporting this view is the fact that it is quite hard to get descriptions of the form 'the man bearing the name *N*' to have narrow scope under metaphysical modals Consider this sentence

(40) The president might not have been the man called 'Havelock'

It is very hard to read 'the president' in (40) as a wide-scope description while reading 'the man called "Havelock"' as a narrow-scope description—in other words it is hard to read the sentence as saying that the actual man who is now the current president might have had a different name So, as we should expect given the conclusions I have reached, metalinguistic descriptions are extremely resistant to getting narrow scope

---

[40] Of course many who hold a descriptive account of names will not be happy with this response because they do not think that it is possible for there to be referring devices without descriptive content They may, however, think that children associate different descriptions with names from those which adult users associate with them

[41] See the literature cited in footnote 38 for discussion of how the metalinguistic view of names might avoid other challenges from Kripke and elsewhere

Of course, in some contexts even metalinguistic descriptions will count as role-type descriptions And in such cases, metalinguistic descriptions will be able to receive narrow-scope interpretations Let us imagine a situation in which it is part of the common ground that there is always one, but not always the same, person bearing a particular name across different situations Consider the name 'M'—the name of the head of the British secret service in *James Bond* 'M' looks like a proper name, but if it is a proper name it is one can get narrow scope with respect to modal operators

(41) John might have become M

The names of superheroes also exhibit this behavior Consider Batman and Superman In different circumstances different individuals may bear the superhero-title [42] Given this fact, it would be appropriate to talk about who might have been Superman or Batman If proper names were just tags attached to particular individuals this behavior would be unexpected We would not expect that the mere presentation of various relevant counterfactual situations across which different individuals lay claim to the same name would allow names to have narrow scope under modal operators So names such as 'M' and 'Superman', unless they are somehow special, or differ in their semantic status from other proper names, provide support for the idea that names are semantically equivalent to definite descriptions, and, thus, in appropriate circumstances, can act as role-type descriptions [43]

The metalinguistic view has many further consequences, however, and we need to see whether they are also supported by our linguistic intuitions about how proper names work For instance, the view entails that names should always show the same potential scope as the definite descriptions that paraphrase them Many have contested this point The following examples, discussed by Barbara Abbott,[44] are supposed to show that names cannot be synonymous with metalinguistic descriptions

(42) Aristotle might not have been Aristotle

(43) Aristotle might not have been the man named 'Aristotle'

---

[42] Apparently there are comic books set in the future in which different individuals are Batman, Superman, and so on

[43] Some, such as Soames in *Beyond Rigidity*, argue that names like these are semantically distinct from other proper names

[44] Abbott, "Definiteness and Proper Names Some Bad News for the Description Theory," *Journal of Semantics*, XIX (2001) 191–201

The usual claim is that (42) has no true reading whereas (43) has a true reading

It is worth pointing out, first of all, that neither sentence easily gets a sensible reading as a metaphysical modal assertion [45] This is evident from the fact that neither (42) nor (43) express the same thing as (44) nor is as obviously true

(44) Aristotle might not have been named 'Aristotle'

This fact, of course, just follows from the earlier observation that particularized descriptions like 'the man named Aristotle' in (43) do not have narrow-scope readings under metaphysical modals A sentence like (43) is not assertable just by virtue of there being a metaphysically possible world where Aristotle is not named 'Aristotle' Rather getting the narrow-scope reading of the description in (43) requires the common ground to include an entire range of relevant possible situations in which the descriptive content is satisfied by different individuals

In certain contexts, a description such as 'the man named "Aristotle"' will be a role-type one For instance, imagine it is commonly known that Greek law ensures that one and only one person is called 'Aristotle' at a single moment of time In this case there may be different relevant possible situations in which different people are uniquely called 'Aristotle' and so the description 'the man named "Aristotle"' acts as a role-type one Then, we might have an interest in

---

[45] Perhaps Abbott's argument is about epistemic modality rather than metaphysical modality In this case, I think the purported difference between (42) and (43) invites a pragmatic explanation This is because, proper names, generally, can get nonrigid narrow-scope readings within epistemic modals A pragmatic explanation of the difference between (42) and (43) could rely on the observation that if a definite noun phrase is repeated twice with the same intonation it is generally taken to have the same referent And indeed if we raise the intonation on the second 'Aristotle' from (43) we get a true reading

(A) Aristotle might not have been ↑ARISTotle

Critics of the descriptivist view would argue that this second use is special and somehow nonliteral However, since what is being debated is whether or not names are descriptions, simply labeling all descriptive uses as nonliteral is question-begging In fact, emphasizing a word does not always make the use nonliteral One might rather think emphasis is meant to show that the two uses of the name are not meant to be taken as coreferential A similar phenomenon occurs with pronouns in these examples

(B) He went to the gym, and ↑HE went to the bathroom
(C) He went to the gym, and he went to the bathroom

So rising intonation can indicate that a word is being used for a different purpose, but not that the use is somehow nonliteral or metalinguistic Sorting out these issues would require more serious work on focus and intonation who would have been called 'Aristotle' if the actual person called 'Aristotle' had not been born Consider this sentence

(45) The person bearing the name 'Aristotle' could have been a sailor

In these circumstances, it seems like it is quite easy to give the description a narrow-scope interpretation

The crucial test for the metalinguistic descriptivist view is whether proper names also allow narrow scope in such circumstances It is unclear what one should say about the sentence containing two proper names, (42), repeated below, when uttered in a context in which a Greek law of this sort is commonly known to be in effect

(42) Aristotle might not have been Aristotle

I think it is perhaps less good than the sentence yielded by replacing the proper names with two definite descriptions

(43) The man called 'Aristotle' might not have been the man called 'Aristotle'

But the difference between the felicity of these two sentences is very subtle, and both of these sentences are rather unnatural A better example of a potential narrow-scope use of a proper name is a variation on (45)

(47) Aristotle could have been a sailor

If there is a Greek law stipulating that there is always one and only one Aristotle at any given time, then (47) seems like it has a reading on which the name gets narrow scope I am not sure whether, with the narrow-scope reading, (47) is less natural than (45) or not In general, I am not sure where the weight of intuitions lies in these cases [46] However, I do not think the intuitions are weighty enough to form the basis of a serious argument against the view that names are semantically equivalent to metalinguistic definite descriptions

We should not despair over the semantics of proper names just because our judgments of critical cases are hazy The messiness of the data is not an obstacle to understanding proper names, it is just another piece of data in its own right The question of whether proper names are particularized descriptions might not have a determinate answer The right hypothesis may be that names are very similar to metalinguistic descriptions, but not exactly the same That is, we may have a conventionally encoded bias towards particularized readings of the descriptive content that names bring with them

---

[46] One has to put aside the question-begging temptation—common in discussions of the modal argument—to label any narrow-scope use of a proper name as somehow special and, hence, irrelevant The important thing to note in this context is that the narrow-scope uses of some definite descriptions, the ones which are usually particularized such as that in (45), also sound slightly odd

What is important to see is that once we restrict our attention to the relevant situations—the cases where names should, on the descriptivist view, get narrow scope—the difference between names and descriptions becomes extremely subtle Altogether the metalinguistic view of proper names does well at predicting what scope proper names will get under modal operators If anything, it does better than standard nondescriptivist views which do not have many resources for explaining the fact that names sometimes do exhibit narrow scope under modal operators

I certainly do not intend this as a serious defense of the metalinguistic view of proper names While the view has its attractions, I am not inclined to think it is correct—if only for the reason that it is hard to explain why, out of the whole space of possible descriptive contents that names might have, names happen to have the metalinguistic content[47] My main point here is just that considerations of scope do not force us to treat proper names as being semantically distinct from definite descriptions

VIII CONCLUSION

I have argued here that definite descriptions and proper names can be treated as expressions designed to pick out entities that satisfy certain presuppositions These presuppositions give rise to different scoping potential in different contexts under modal operators and adverbs of quantification I think this position should be the default one, since it respects the basic intuition that singular terms including both descriptions and names share a deep semantic commonality Since this view also provides a better analysis of the interaction of modals with definite descriptions and proper names, it has much to recommend it

I hope my arguments here can help clarify the philosophical discussion of singular terms Much of what interests philosophers in this area is the (apparent) capacity of singular terms to directly refer to objects and individuals Theorists have tended to focus on the semantics of proper names in their discussions of such direct reference However, if the semantics of names does not differ substantially from that of descriptions then direct reference is more likely to be an aspect of language use rather than a general semantic property of a category of expressions [48]

---

[47] I am inclined to think linguistic usage (in speech or in the head) does not determinately fix the descriptive content of names Thus, there is no fact of the matter about what the descriptive content of names is

DANIEL ROTHSCHILD

Columbia University

---

[48] Of course, direct reference theorists also argue that other terms such as demonstratives and indexicals are directly referential I would hope the line of argument developed here might be expanded to encompass indexicals and demonstratives also

NEW BOOKS ANTHOLOGIES

HART, KEVIN, and GEOFFREY H HARTMAN, eds *The Power of Contestation Perspectives on Maurice Blanchot* Baltimore, MD Johns Hopkins, 2004 ix, 222 p Cloth $49 95

HART, KEVIN, and BARBARA WALL, eds *The Experience of God A Postmodern Response* New York Fordham, 2005 xii, 259 p Cloth $ 75 00, paper $25 00

HASLANGER, SALLY, and CHARLOTTE WITT, eds *Adoption Matters Philosophical and Feminist Essays* Ithaca, NY Cornell, 2005 xiii, 320 p Cloth $49 95, paper $22 95

HASLANGER, SALLY, and ROXANNE MARIE KURTZ, eds *Persistence Contemporary Readings* Cambridge MIT, 2007 ix, 483 p Paper $36 00

HELDKE, LISA, KERRI MOMMER, and CYNTHIA PINEO, eds *The Atkins Diet and Philosophy* Chicago Open Court, 2005 xv, 249 p Paper $17 95

HELMER, CHRISTINE, and CHRISTOF LANDMESSER, eds *One Scripture or Many? Canon from Biblical, Theological, and Philosophical Perspectives* New York Oxford, 2004 xi, 249 p Cloth $95 00

HENDRICKS, VINCENT, and JOHN SYMONS, eds *Formal Philosophy* New York Automatic Press, 2005 v, 246 p Cloth $71 50, paper $22 95

HERMAND, JOST, and GERHARD RICHTER, eds *Sound Figures of Modernity German Music and Philosophy* Madison Wisconsin UP, 2006 vii, 267 p Cloth $39 95

HERWITZ, DANIEL, and MICHAEL KELLY, eds *Art, Action, History Engagements with Arthur C Danto* New York Columbia, 2007 xii, 238 p Cloth $45 00

HETHERINGTON, STEPHEN, ed *Epistemology Futures* New York Oxford, 2006 x, 241 p Cloth $99 00, paper $35 00

HOLDER, JOHN J , ed *Early Buddhist Discourses* Indianapolis Hackett, 2006 xxiii, 216 p Cloth $37 95, paper $12 95

HOLTUG, NILS, and KASPER LIPPERT-RASMUSSEN, eds *Egalitarianism New Essays on the Nature and Value of Equality* New York Oxford, 2007 xi, 339 p Cloth $74 00

HOLYOAK, KEITH J , and ROBERT G MORRISON, eds *The Cambridge Handbook of Thinking and Reasoning* New York Cambridge, 2005 xiv, 858 p Cloth $120 00, paper $60 00

HONDERICH, TED, ed *The Oxford Companion to Philosophy New Edition* New York Oxford, 2005 xix, 1056 p Cloth $60 00

HORGAN, TERRY, and MARK TIMMONS, eds *Metaethics after Moore* New York Oxford, 2006 xi, 397 p Cloth $99 00, paper $35 00

HORNSBY, JENNIFER, and GUY LONGWORTH, eds *Reading Philosophy of Language Selected Texts with Interactive Commentary* Malden, MA Blackwell, 2006 xi, 324 p Cloth $64 95, paper $29 95

HÖSLE, VITTORIO, and CHRISTIAN ILLIES, eds *Darwinism and Philosophy* Notre Dame University Press, 2005 400 p Cloth $70 00, paper $35 00

HURLEY, SUSAN, and NICK CHATER, eds *Perspectives on Imitation From Neuroscience to Social Science, Volumes 1 and 2* Cambridge MIT, 2005 xi, 437, xi, 547 p Cloth $90 00, paper $35 00

HYDE, MICHAEL J , ed *The "Ethos" of Rhetoric* Columbia South Carolina UP, 2004 xxviii, 231 p Cloth $39 95

HYLAND, DREW, and JOHN PANTELEIMON MANOUSSAKIS, eds *Heidegger and the Greeks Interpretive Essays* Bloomington Indiana UP, 2007 viii, 194 p Cloth $65 00, paper $24 95

IRWIN, WILLIAM, ed *More Matrix and Philosophy* Chicago Open Court, 2005 xi, 220 p Paper $17 95

JACKSON, FRANK, and MICHAEL SMITH, eds *The Oxford Handbook of Contemporary Philosophy* New York Oxford, 2006 xii, 904 p Cloth $99 00

JAHN, BEATE, ed *Classical Theory in International Relations* New York Cambridge, 2006 xv, 309 p Cloth $80 00, paper $32 99

JAROSZYŃSKI, PIOTR, ed *Science in Culture* New York Rodopi, 2006 xx, 314 p Paper $88 00

JENKINS, JOYCE, JENNIFER WHITING, and CHRISTOPHER WILLIAMS, eds *Persons and Passions Essays in Honor of Annette Baier* Notre Dame University Press, 2005 viii, 368 p Cloth $53 00

JUDSON, LINDSAY, and VASSILIS KARASMANIS, eds *Remembering Socrates Philosophical Essays* New York Oxford, 2006 viii, 207 p Cloth $74 00

KALDERON, MARK ELI, ed *Fictionalism in Metaphysics* New York Oxford, 2005 x, 354 p Cloth $95 00, paper $35 00

KANARIS, JIM, and MARK J DOORLEY, eds *In Deference to the Other* Albany SUNY, 2004 xiii, 187 p Cloth $35 00

KELLERT, STEPHEN H , HELEN E LONGINO, and C KENNETH WATERS, eds *Scientific Pluralism* Minneapolis Minnesota UP, 2006 xxix, 248 p Cloth $50 00

KHALIDI, MUHAMMAD ALI, ed *Medieval Islamic Philosophical Writings* New York Cambridge, 2005 xlviii, 186 p Cloth $65 00, paper $25 99

KIM, JAEGWON, ed *Philosophy of Mind, 2nd Edition* Boulder, CO Westview, 2005 xii, 338 p Cloth $59 00, paper $29 00

KINCAID, HAROLD, JOHN DUPREE, and ALISON WYLIE, eds *Value-Free Science? Ideals and Illusions* New York Oxford, 2007 xiv, 241 p Cloth $65 00

KOERTGE, NORETTA, ed *Scientific Values and Civic Virtues* New York Oxford, 2005 x, 245 p Cloth $70 00, paper $24 95

KRAUT, RICHARD, ed *The Blackwell Guide to Aristotle's Nicomachean Ethics* Malden, MA Blackwell, 2006 x, 384 p Cloth $74 95, paper $34 95

KRAUT, RICHARD, and STEVEN SKULTETY, eds *Aristotle's Politics Critical Essays* New York Rowman and Littlefield, 2005 xviii, 256 p Cloth $75 00, paper $27 95

KRITZMAN, LAWRENCE D , ed *The Columbia History of Twentieth-Century French Thought* New York Columbia, 2006 xxv, 787 p Cloth $75 00

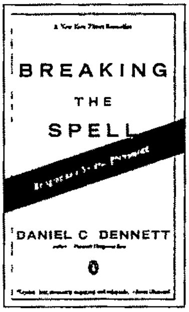
KUHSE, HELGA, and PETER SINGER, eds *Bioethics An Anthology* Malden, MA Blackwell, 2006 xvii, 738 p Cloth $99 95

KUKLA, REBECCA, ed *Aesthetics and Cognition in Kant's Critical Philosophy* New York Cambridge, 2006 xiii, 309 p Cloth $75 00

KUPER, ANDREW, ed *Global Responsibilities Who Must Deliver Human Rights?* New York Routledge, 2005 xxii, 283 p Paper $26 95

KYMLICKA, WILL, and BAOGANG HE, eds *Multiculturalism in Asia* New York Oxford, 2005 xi, 364 p Cloth $99 00, paper $35 00

NEW FROM PENGUIN GROUP (USA)



DANIEL C. DENNETT
BREAKING THE SPELL
Religion as a Natural Phenomenon
Penguin 464 pp 0-14-303833-8 $16.00

MICHAEL KAPLAN & ELLEN KAPLAN
CHANCES ARE
Adventures in Probability
A journey through history, mathematics, and philosophy, charting humanity's struggle against randomness
Penguin 336 pp 0-14-303834-6 $15.00

MARQUIS DE SADE
PHILOSOPHY IN THE BOUDOIR
Introduction by Francine du Plessix Gray
Translated by Joachim Neugroschel
Cover by Tomer Hanuka
Deluxe edition with French flaps
Penguin Classics 192 pp 0-14-303901-6 $14.00

MARCUS AURELIUS
MEDITATIONS
Newly Translated and Edited with Notes by Martin Hammond. Introduction by Diskin Clay
Penguin Classics 304 pp 0-14-044933-7 $10.00

JACOB NEEDLEMAN
WHY CAN'T WE BE GOOD?
"[Needleman is] an immensely learned man who is unembarrassed by the big questions that many of his fellow philosophers prefer to dodge with an ironically raised eyebrow"—*Chicago Tribune*
Tarcher 256 pp 1-58542-541-9 $24.95

GARY LACHMAN
RUDOLF STEINER
An Introduction to His Life and Work
Tarcher 304 pp 1-58542-543-5 $16.95

JUDITH P ZINSSER
LA DAME D'ESPRIT
A Biography of the Marquise Du Châtelet
Viking 400 pp 0-670-03800-8 $24.95

THE TIBETAN BOOK OF THE DEAD
Translated by Gyurme Dorje, Edited by Graham Coleman with Thupten Jinpa, Introductory Commentary by His Holiness The Dalai Lama
Deluxe edition with French flaps
Penguin Classics 592 pp 0-14-310494-2 $20.00

PENGUIN GROUP (USA) www.penguingroup.com/academic
Academic Marketing Department 375 Hudson Street NY, NY 10014

New from Stanford University Press

Reflections on Literature and Culture

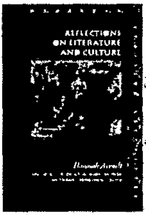
HANNAH ARENDT

Edited and with an Introduction by SUSANNAH YOUNG-AH GOTTLIEB

Including previously unpublished and untranslated material drawn from half a century of engagement with the works of European and American authors, poets, journalists, and literary critics, this is the first volume in any language that collects Hannah Arendt's remarkable series of essays and notes on literary figures and cultural questions

Meridian Crossing Aesthetics

$24 95 paper $65 00 cloth



AVAILABLE IN MARCH 2007

Speaking of Freedom

Philosophy, Politics, and the Struggle for Liberation

DIANE ENNS

"This book manages to convince the reader that speaking of freedom really matters Enns focuses judiciously and expertly on the influential theories of freedom running from existentialism and phenomenology to deconstruction and post-structuralism She brings a real sense of engagement to her study which makes the arguments lift off the page "

—Richard Kearney, Boston College

$55 00 cloth

Stanford University Press

800 621 2736 www sup org

NEW FROM LEXINGTON BOOKS . . .

Puzzles & Perplexities

Collected Essays, Second Edition

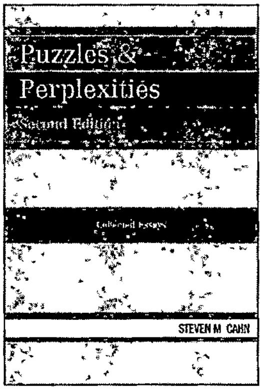
By Steven M Cahn

"Cahn is a master of the philosophical essay In each chapter he confronts a perennial issue, such as God, free will, art, and social justice, and provides an original and provocative philosophical discussion The chapters combine crisp thinking, elegant writing, and careful, imaginative analysis These essays are models of philosophical writing"

—Robert Talisse, Teaching Philosophy

"Not only Cahn's lucidity but his philosophical passion and honesty shine through these essays He is both scrupulously fair to opposing positions and straightforward in advocating his own, whether popular or not The book will engage students and scholars alike I found it a pleasure to read"

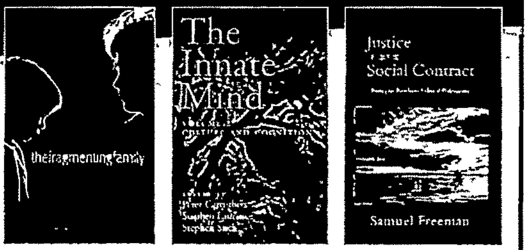
—Alan H. Goldman, Newsletter on Teaching Philosophy



Available June 2007
176 pages, $29.95 paper
ISBN 0-7391-2116-2
ISBN-13 978-0-7391-2116-0

Order online at www.lexingtonbooks.com • 1.800.462.6420

New and Recent Texts
from
OXFORD



The Fragmenting Family
BRENDA ALMOND
Brenda Almond throws down a timely challenge to liberal consensus about personal relationships She maintains that the traditional family is fragmenting in Western societies, and that this fragmentation is a cause of serious social problems She urges that we reconsider our attitudes to sex and reproduction in order to strengthen our most important social institution, the family
2006 200 pp $24 95

New in Paperback!
Conversations on Consciousness
What the Best Minds Think about the Brain, Free Will, and What It Means to Be Human
SUSAN BLACKMORE
"Succeeds in providing a very brief survey of the multitude of positions occupied by thinkers in this area The often quirky personalities and mannerisms of the interviewees shine through the text. Blackmore herself comes across as spunky and clever, and the probing follow-up questions she occasionally asks prevent the interviews from seeming too repetitive and boring" —*Nature*
2006 (paper 2007) 288 pp, 30 halftone paper $15 95 cloth $23 00

Justice and the Social Contract
Essays on Rawlsian Political Philosophy
SAMUEL FREEMAN
"Freeman is the leading authority on the thought and writing of John Rawls, and Rawls was the leading political and social philosopher of the twentieth century Freeman's clear, careful, and deeply informed studies in these essays offer important insight about basic questions of interpretation and justification"
—Thomas Nagel, New York University
2006 352 pp $55 00

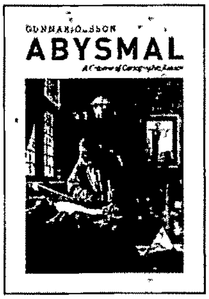
The Innate Mind
Volume 2 Culture and Cognition
Edited by PETER CARRUTHERS, STEPHEN LAURENCE, and STEPHEN STICH
"If you are impressed by our species' cross-cultural variation, but mystified by how it emerges, then read this book. If you are unimpressed by our cross-cultural variation, but moved by our universality, then read this book You will learn that of all the species that have existed on earth, we are truly bizarre, at once both clones of each other and yet remarkably different"
—Marc Hauser, Harvard College Professor of Psychology and Organismic and Evolutionary Biology, and author of *Wild Minds* and *Moral Minds*
(Evolution and Cognition Series)
2007 368 pp, 37 line illus paper $45 00 cloth $99 00

New in Paperback!
God, the Devil, and Darwin
A Critique of Intelligent Design Theory
NIALL SHANKS
Foreword by RICHARD DAWKINS
"[A] cogent and well-argued alarum Shanks deftly skewers the scientific pretensions of intelligent design creationists"—*Science*
Backed by intellectuals at respectable universities, Intelligent Design theory (ID) proposes an alternative to accepted accounts of evolutionary theory that life is so complex, and that the universe is so fine-tuned for the appearance of life, that the only plausible explanation is the existence of an intelligent designer For many ID theorists, the designer is taken to be the god of Christianity Niall Shanks has written the first accessible introduction to, and critique of, this controversial new intellectual movement
2004 (paper February 2007) 296 pp, 3 b/w illus paper $14 95 cloth $32 00

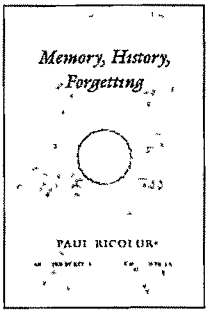
OXFORD
UNIVERSITY PRESS

Prices are subject to change and apply only in the US.
To order, please call 1-800-451-7556.
In Canada, call 1-800-387-8020.
Visit our web site at www.oup.com/us/social work.

PHILOSOPHY
new from Chicago



ABYSMAL
A Critique of Cartographic Reason
Gunnar Olsson
"*Abysmal* is a massive undertaking that captivates the reader at every twist and turn Complex ideas are developed in relationship to one another, refracting across many primary texts of Western thought To read each chapter is to experience a heady, sparkling flow of thought-in-action "—John Pickles, University of North Carolina at Chapel Hill
CLOTH $40 00



Now in paperback
MEMORY, HISTORY, FORGETTING
Paul Ricoeur
Translated by Kathleen Blamey and David Pellauer
"*Memory, History, Forgetting* is an exceedingly serious study serious in the sense of being thoughtful, thorough, with a good sense of what is important "
—*Times Literary Supplement*

"Ricoeur writes the best kind of philosophy—critical, economical, and clear "—*New York Times Book Review*
PAPER $25 00

THE OBJECTIVE EYE
Color, Form, and Reality in the Theory of Art
John Hyman
"One of the many virtues of Hyman's book is that it provides a sustained, rigorous and devastating critique of subjectivism, and a carefully nuanced defense of his own version of objectivism "
—*London Review of Books*
PAPER $35 00

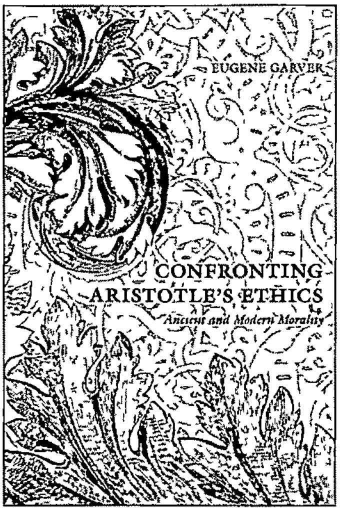
CSLI PUBLICATIONS

New from CSLI
EUROPEAN REVIEW OF PHILOSOPHY, VOLUME 6
The Structure of Nonconceptual Content
Edited by Christine van Geen and Frédérique de Vignemont
The European Review of Philosophy, Volume 6 brings together contributions from both conceptualists and nonconceptualists to shed new light on issues of interest to cognitive scientists and philosophers of mind
PAPER $25 00

THE UNIVERSITY OF CHICAGO PRESS • www press uchicago edu

PHILOSOPHY

from Chicago



EUGENE GARVER

CONFRONTING ARISTOTLE'S ETHICS

Ancient and Modern Morality

CONFRONTING ARISTOTLE'S ETHICS

Ancient and Modern Morality

Eugene Garver

"Eugene Garver has a clear and unified vision of Aristotle's project in ethics and, in particular, of the importance of Aristotelian politics and metaphysics in this understanding His is an entirely new and refreshing approach, one that challenges many of the mainstream accounts of Aristotle "—Charles Young, Claremont Graduate School

CLOTH $49 00

THE DERRIDA-HABERMAS READER

Edited by Lasse Thomassen

"This is a most welcome volume It provides the reader not only with a wonderful overview of the deep philosophical differences that divide Derrida and Habermas, but also with some sense of their common hopes as Europeans and cosmopolitan citizens "—Stephen K White, University of Virginia

PAPER $29 00

DARWINIAN REDUCTIONISM

Or, How to Stop Worrying and Love Molecular Biology

Alex Rosenberg

"In this stimulating new book, Rosenberg's Darwinian reductionism denies genetic determinism and it has no problem with multiple realizability It captures what scientific materialism should have been after all along "—Elliot Sober, University of Wisconsin

CLOTH $40 00

FASHION

A Philosophy

Lars Svendsen

An engaging analysis of fashion and the motivations behind it, *Fashion* analyzes this subject as both a historical phenomenon and a philosophy of aesthetics

Distributed for Reaktion Books

PAPER $24 95

Now in Paperback

TIME, CREATION AND THE CONTINUUM

Theories in Antiquity and the Early Middle Ages

Richard Sorabji

PAPER $22 50

THE UNIVERSITY OF CHICAGO PRESS • www press uchicago edu

CAMBRIDGE

Outstanding Philosophy Titles

An Introduction to Political Philosophy
Colin Bird
Cambridge Introductions to Philosophy
$80.00 Hb 978-0-521-83625-8
$24.99 Pb 978-0-521-54482-5
322 pp

The Cambridge History of Literary Criticism
Volume 5 - Romanticism
Edited by Marshall Brown
$140.00 Hb 978-0-521-30010-0
$34.99 Pb 978-0-521-31721-4
512 pp

Hume: *An Enquiry Concerning Human Understanding*
And Other Writings
Edited by Stephen Buckle
Cambridge Texts in the History of Philosophy
$50.00 Hb 978-0-521-84340-9
$22.99 Pb 978-0-521-60403-1
278 pp

Hunting Causes and Using Them
Approaches in Philosophy and Economics
Nancy Cartwright
$85.00 Hb 978-0-521-86081-9
$29.99 Pb 978-0-521-67798-1
280 pp

Neurophilosophy at Work
Paul Churchland
$75.00 Hb 978-0-521-86472-5
$24.99 Pb 978-0-521-69200-7
262 pp

Hume: *Dialogues Concerning Natural Religion*
And Other Writings
Edited by Dorothy Coleman
Cambridge Texts in the History of Philosophy
$55.00 Hb 978-0-521-84260-0
$19.99 Pb 978-0-521-60359-1
310 pp

Explaining Social Behavior
More Nuts and Bolts for the Social Sciences
Jon Elster
$85.00 Hb 978-0-521-77179-5
$27.99 Pb 978-0-521-77744-5
496 pp

Heidegger's *Being and Time*
An Introduction
Paul Gorner
Cambridge Introductions to Key Philosophical Texts
$80.00 Hb 978-0-521-83322-6
$23.99 Pb 978-0-521-54072-8
204 pp

A History of Natural Philosophy
From the Ancient World to the Nineteenth Century
Edward Grant
$70.00 Hb 978-0-521-86931-7
$24.99 Pb 978-0-521-68957-1
376 pp

Spinoza: *Theological-Political Treatise*
Edited by Jonathan Israel and Michael Silverthorne
Cambridge Texts in the History of Philosophy
$75.00 Hb 978-0-521-82411-8
$19.99 Pb 978-0-521-53097-2
330 pp

Wordsworth's Philosophic Song
Simon Jarvis
Cambridge Studies in Romanticism
$85.00 Hb 978-0-521-86268-4
284 pp

Kant and the Power of Imagination
Jane Kneller
$85.00 Hb 978-0-521-85143-5
180 pp

An Introduction to the Theory of Knowledge
Noah Lemos
Cambridge Introductions to Philosophy
$85.00 Hb 978-0-521-84213-6
$24.99 Pb 978-0-521-60309-6
242 pp

Rethinking Informed Consent in Bioethics
Neil C. Manson and Onora O'Neill
$85.00 Hb 978-0-521-87458-8
$32.99 Pb 978-0-521-69747-7
226 pp

War Crimes and Just War
Larry May
$80.00 Hb 978-0-521-87114-3
$29.99 Pb 978-0-521-69153-6
359 pp

Leibniz's Metaphysics
Its Origins and Development
Christia Mercer
$120.00 Hb 978-0-521-40301-6
$55.00 Pb 978-0-521-02992-6
542 pp

Prices subject to change.

www.cambridge.org/us/philosophy
1-800-872-7423

CAMBRIDGE
UNIVERSITY PRESS

CAMBRIDGE

from Cambridge

Discovering Levinas
Michael L Morgan
$90 00 Hb 978-0-521-87259-1
490 pp

An Introduction to the Philosophy of Language
Michael Morris
Cambridge Introductions to Philosophy
$75 00 Hb 978-0-521-84215 0
$29 99 Pb 978-0-521-60311-9
336 pp

The Cambridge Companion to Locke's *Essay Concerning Human Understanding*
Edited by Lex Newman
Cambridge Companions to Philosophy
$80 00 Hb 978 0-521-83433-9
$29 99 Pb 978-0-521-54225 8
502 pp

The Price of Peace
Just War in the Twenty-First Century
Edited by Charles Reed and David Ryall
Foreword by Richard Dannatt
$85 00 Hb 978-0-521-86051-2
$29 99 Pb 978-0-521-67785-1
358 pp

A Natural History of Pragmatism
The Fact of Feeling from Jonathan Edwards to Gertrude Stein
Joan Richardson
Cambridge Studies in American Literature and Culture
$85 00 Hb 978-0-521-83748-4
$28 99 Pb 978 0-521-69450 6
346 pp

4 Volume Set!

Richard Rorty. *Philosophical Papers*
$120 00 4 Pb books
978-0 521-70152-5

Philosophy as Cultural Politics
Volume 4
Philosophical Papers
Richard Rorty
$80 00 Hb 978-0-521-87544-8
$22 99 Pb 978-0-521-69835-1
218 pp

Group Rationality in Scientific Research
Husain Sarkar
$80 00 Hb 978-0-521-87113-6
300 pp

Ayn Rand's Normative Ethics
The Virtuous Egoist
Tara Smith
$80 00 Hb 978-0 521-86050-5
$25 99 Pb 978 0-521-70546-2
328 pp

Roman Monarchy and the Renaissance Prince
Peter Stacey
Ideas in Context
$90 00 Hb 978 0-521-86989-8
360 pp

Appearances of the Good
An Essay on the Nature of Practical Reason
Sergio Tenenbaum
$80 00 Hb 978-0-521-83783-5
326 pp

Fallacies and Argument Appraisal
Christopher W Tindale
Critical Reasoning and Argumentation
$75 00 Hb 978-0-521-84208-2
$24 99 Pb 978-0-521-60306-5
236 pp

J S Mill's Political Thought
A Bicentennial Reassessment
Edited by Nadia Urbinati and Alex Zakaras
$85 00 Hb 978-0-521-86020-8
$27 99 Pb 978-0-521-67756-1
400 pp

Recognition and Power
Axel Honneth and the Tradition of Critical Social Theory
Bert van den Brink and David Owen
$85 00 Hb 978-0-521-86445-9
416 pp

Liberalism, Perfectionism and Restraint
Steven Wall
$70 00 Hb 978 0-521-62411-4
$32 99 Pb 978-0-521-03542-2
256 pp

Locke A Biography
R S Woolhouse
$39 99 Hb 978-0-521-81786-8
558 pp

Deliberative Democracy and the Institutions of Judicial Review
Christopher F Zurn
$85 00 Hb 978-0-521-86734-4
374 pp

Prices subject to change

www cambridge org/us/philosophy
1-800-872-7423

CAMBRIDGE
UNIVERSITY PRESS

ANNOUNCING:

THE JOURNAL OF PHILOSOPHY

ONLINE

The Journal of Philosophy has added
a print and online subscription option for institutions

SUBSCRIPTION OPTIONS

Libraries and Institutions (print and online)	$180.00
Libraries and Institutions (print only)	$100.00
Individuals	$45.00
Students, retired/unemployed philosophers	$20.00
Postage outside the U.S.	$15.00

Please consult our website at www.journalofphilosophy.org
for more details

◆

Bailan
30.8.07

THE JOURNAL OF PHILOSOPHY

VOLUME CIV, NUMBER 3
MARCH 2007

page
109 *Collective Intentions and Team Agency* Natalie Gold and Robert Sugden
138 *Collectivity and Circularity* Björn Petersson
157 *Knowledge and Evidence* Joseph Mendola
161 NEW BOOKS ANTHOLOGIES

Published by **The Journal of Philosophy, Inc.**

THE JOURNAL OF PHILOSOPHY

FOUNDED BY FREDERICK J E WOODBRIDGE AND WENDELL T BUSH

Purpose To publish philosophical articles of current interest and encourage the interchange of ideas, especially the exploration of the borderline between philosophy and other disciplines

Editors Bernard Berofsky, Akeel Bilgrami, John Collins, Arthur C Danto, Kent Greenawalt, Patricia Kitcher, Philip Kitcher, Isaac Levi, Wolfgang Mann, Mary Mothersill, Frederick Neuhouser, Christopher Peacocke, Carol Rovane, Achille C Varzi *Consulting Editors* David Albert, James T Higginbotham, Robert May, Charles D Parsons, Wilfried Sieg *Managing Editor* John Smylie

THE JOURNAL OF PHILOSOPHY is owned and published by the Journal of Philosophy, Inc *President,* Arthur C Danto, *Vice President,* Akeel Bilgrami, *Secretary,* Daniel Shapiro, *Treasurer,* Barbara Gimbel, *Other Trustees* Lee Bollinger, Leigh S Cauman, Kent Greenawalt, Michael J Mooney, Lynn Nesbit

All communications to the Editors and Trustees and all manuscripts may be sent to John Smylie, Managing Editor, Mail Code 4972, 1150 Amsterdam Avenue, Columbia University, New York, New York 10027 FAX (212) 932-3721
You may also visit our website at www journalofphilosophy org

THE JOURNAL OF PHILOSOPHY

2007

SUBSCRIPTIONS (12 issues)

Individuals	$45 00
Libraries and Institutions	$100 00
Libraries and Institutions (print and online)	$180 00
Students, retired/unemployed philosophers	$20 00
Postage outside the U S	$15 00

Payments only in U S currency on a U S bank All back volumes and separate issues available back to 1904 Please inquire for advertising rates, ad space is limited, so ad reservations are required

Published monthly as of January 1977, typeset by Dartmouth Journal Services, Waterbury, VT, and printed by The Sheridan Press, Hanover, PA

All communication about subscriptions and advertisements may be sent to Pamela Ward, Business Manager, Mail Code 4972, 1150 Amsterdam Avenue, Columbia University, New York, NY 10027 (212) 866-1742

All materials published in *The Journal of Philosophy* are copyrighted by the *Journal* and are protected by U S copyright law All rights reserved For more information about our copyright policies, please consult our January issue or contact us

The Journal of Philosophy (ISSN 0022-362X) is published monthly by The Journal of Philosophy, Inc , MC4972, Columbia University, 1150 Amsterdam Avenue, New York, NY 10027 Periodicals postage paid at New York, NY, and other mailing offices
POSTMASTER Please send address changes to *The Journal of Philosophy* at MC4972, Columbia University, 1150 Amsterdam Avenue, New York, NY 10027

© Copyright 2007 by the Journal of Philosophy, Inc ISSN 0022-362X

University

THE JOURNAL OF PHILOSOPHY
VOLUME CIV, NO 3, MARCH 2007

COLLECTIVE INTENTIONS AND TEAM AGENCY*

Collective intentions are those intentions associated with joint actions The paradigm cases include two people singing a duet, painting a house, pushing a car, and taking a walk together In these situations, there is a sense in which "we" intend the joint action, as well as a sense in which "I" intend my part in it For instance, if we go to Edinburgh together and, as a part of this, I check the train timetable and you buy the tickets, then there is a sense in which our actions result from "our" intention to go to Edinburgh This collective intention is not reducible to the summation of our individual intentions and, as such, is unlike the case where two agents happen to do the same thing independently For instance, if I plan to go to Edinburgh and coincidentally meet you on the train, there is a use of the word 'we' in which it may be that "we went to Edinburgh" because "I went to Edinburgh" and "you went to Edinburgh," but this is not a joint action and the intentions behind it are not collective intentions Analyses of collective intentions seek to elucidate the features that are peculiar to the joint action case and to explain how collective intentions are connected to individual intentions

The literature on collective intentions is exemplified by the work of Raimo Tuomela and Kaarlo Miller,[1] John Searle,[2] and Michael Bratman [3] A general problem for these accounts is how to differentiate collective intentions from the mutually-consistent individual intentions that lie behind Nash equilibrium behavior in games In game theory, the

---

* We would like to thank Nick Bardsley, Michael Bratman, Margaret Gilbert, Philip Pettit, and Wlodek Rabinowicz for helpful discussions

[1] Tuomela and Miller, "We-intentions," *Philosophical Studies*, LIII (1988) 367–89

[2] Searle, "Collective Intentions and Actions," in Philip R. Cohen, Jerry Morgan, Martha E Pollack, eds , *Intentions in Communication* (Cambridge MIT, 1990), pp 401–15

[3] Bratman, "Shared Cooperative Activity," *Philosophical Review*, CI (1992) 327–41, and "Shared Intention," *Ethics*, CIV (1993) 97–113

0022-362X/07/0403/109-37

behavior of the players of a noncooperative game is defined to be in Nash equilibrium if each player's chosen strategy is optimal for her, given the strategies chosen by the others. Nash equilibrium is usually interpreted as a situation in which each player acts as a rational individual agent, holding true beliefs about the actions of the others. It is clear that not all Nash equilibria are joint actions. However, the core analyses provided by Tuomela and Miller, Searle, and Bratman seem to imply that all Nash equilibrium situations are instances of collective intentions. Cases in which Nash equilibria are not joint actions are excluded only by stipulation or by the addition of further conditions which are just as problematic as the original concept of collective intention. Tuomela and Miller stipulate that the definition of collective intention includes the condition that the action is joint, Searle that collective intentions involve cooperation in pursuit of collective goals. This amounts to saying that the special feature of collective intentions that distinguishes them from the intentions behind Nash equilibrium behavior is that they are associated with cooperative activity, but this is something that we already knew prior to the analysis. Bratman adds conditions which require each agent to be responsive to the behavior of the other as the joint action proceeds and if unexpected problems occur, but these conditions are stated only informally, and rely on a pre-analytic understanding of the nature of cooperative activity.[4]

We shall argue that these problems arise because, in the literature of collective intentions, the focus is on the properties of these intentions as mental states, rather than on the mental processes by which they were formed. Thus, anything that is distinctive to cooperative activity has to be represented as a distinctive feature of the corresponding intentions. An alternative approach is to analyze the practical reasoning by which individual agents choose to engage in cooperative activity and then to consider whether this reasoning has special features which lead to collective intentions.

A starting point for such an analysis can be found in a body of decision-theoretic literature on *team agency*. This seeks to extend standard game theory, where each individual asks separately "What should *I* do?" to allow teams of individuals to count as agents and for players to ask the question "What should *we* do?" This leads to *team reasoning*, a distinctive mode of reasoning that is used by members of teams, and which may result in cooperative actions. When an agent deliberates about what she ought to do, the result of her reasoning is an intention An intention is interposed between reasoning and action, so it is natural to treat the intentions that result from team reasoning as collective intentions The concept of intention seems to imply not reasoning again, or at least that the result of reasoning is a default plan which is revised only under special circumstances So, if the distinctive feature of collective intentions is to be found in the reasoning by which they were formed, then an analysis that focuses on the intentions themselves will miss the feature that makes collective intentions collective In this paper, we will show how team reasoning fills the gaps in the collective intention literature and provide a framework in which various theories of group agency can be compared

---

[4] Bratman, "Shared Cooperative Activity."

I COLLECTIVE INTENTIONS AND NASH EQUILIBRIUM

An early analysis of collective intentionality is the work of Tuomela and Miller (*op cit*) The essential features of this analysis can be presented as follows for the case of a two member group, whose members are P1 and P2 Consider some "joint social action" A which comprises actions A_1 and A_2 for the respective individuals According to Tuomela and Miller, P1 has a *we-intention* with respect to A if (i) P1 intends to do A_1, (ii) P1 believes that P2 will do A_2, (iii) P1 believes that P2 believes that P1 will do A_1, and so on (*op cit*, p 375) This analysis reduces we-intentions to individual intentions and a network of mutual beliefs

An unsatisfactory feature of this analysis is that it seems to treat every Nash equilibrium as a case of collective intentionality In a Nash equilibrium, each individual's action is a best response to her true beliefs about the other's action Since these are intentional actions, this is equivalent to saying that each individual's intention is adapted to her true beliefs about the actions of the other For example, consider the version of the Hawk-Dove game shown in Figure 1 (following)

As an example of such a game, think of two individuals in a state of nature who come into conflict over some valuable resource To play *dove* is to offer to share the resource but to back down if the other attempts to take it all, to play *hawk* is to demand the whole resource, backed by a readiness to fight for it We assume that fighting is costly for both parties, and that the utility value of a half share of the resource is greater than half of the utility value of the whole

This game has two pure-strategy Nash equilibria (*hawk, dove*) and (*dove, hawk*) Consider the first of these Suppose it is common knowledge between P1 and P2 that, in interactions like this, the player in the position of P1 almost always chooses *hawk* and the one in the position of P2 almost always chooses *dove* Expecting P2 to play *dove*, P1 forms the intention to play *hawk* Expecting P1 to play *hawk*, P2

		P2	
		dove	*hawk*
P1	*dove*	**2, 2**	**0, 3**
	hawk	**3, 0**	**−5, −5**

Figure 1 Hawk-Dove

forms the intention to play *dove* Given all this, does each player have a we-intention with respect to the pair of strategies (*hawk, dove*)? On Tuomela and Miller's analysis, it seems that they do

The problem becomes even more acute when we consider the Prisoner's Dilemma, a two-person version of which is shown in Figure 2

As an example of such a game, think of two individuals who must each decide how much of a valuable resource to take This resource renews itself each season but the amount available next season will depend on how much is left at the end of this one (think of fish in the sea) Payoffs reflect the amount of the resource that each player gets over two seasons To play *cooperate* is to take a moderate amount of the resource this season, to play *defect* is to take a large amount The rate of depletion is such that, if only one player takes a large amount then the increased amount she gets this season outweighs the decrease next season, but if one player takes a large amount this season then the other player does better by also taking a large amount now before the resource becomes depleted For each player, whatever the other player does, her best move is to play *defect* However, the outcome in which both players *defect* and the resource is severely depleted is worse for each player than that in which they both *cooperate* and conserve enough of the resource to renew itself for next season

The unique Nash equilibrium of this game is (*defect, defect*) So, in the Nash equilibrium, P1 will form the intention to play *defect* Because playing *defect* is a dominant strategy, if there is common knowledge of rationality, P1 will expect P2 to do likewise (though it is in her best interests to form the intention to *defect* regardless of her expectation about P2's action) So, again, the Nash equilibrium behavior seems to fall under Tuomela and Miller's analysis of collective intentions But, intuitively, the exploitation of the resource is not a joint action It is not true to say that each player intends that "we" exploit the common resource Conversely, the non-Nash equilibrium cooperative practice of conserving the resource does involve a collective intention

Although Tuomela and Miller's core analysis seems to include the intentions behind the Nash equilibria in these games, it comes with various qualifications In particular, it applies only to "joint social

		P2	
		cooperate	***defect***
P1	***cooperate***	**3, 3**	**1, 4**
	defect	**4, 1**	**2, 2**

Figure 2 The Prisoner's Dilemma

actions," defined as "situations in which some agents act together, usually or often with the purpose of achieving some joint goal" (*op cit*, p 367), this goal is "normally (but not necessarily) the goal to perform the total action [in our notation, A]" (*op cit*, p 370) Tuomela and Miller also add a condition to the effect that when P1 performs A_1, "he does it in order for the participating agents to succeed in doing [A]" (*op cit*, p 376) These conditions may rule out some dominant-strategy Nash equilibria as cases of collective intention For example, in the case of the Prisoner's Dilemma, one might deny that P1 plays *defect* in order for P1 and P2 to succeed in playing (*defect, defect*), rather, P1 plays *defect* because that is best for him, irrespective of what P2 does Possibly, these conditions are also intended to exclude cases like the Hawk-Dove example, but if so, how these cases are excluded remains obscure

Searle also criticizes Tuomela and Miller's analysis for including cases where there is no collective intentionality (*op cit*) He proposes that we-intentions cannot be reduced to such combinations of I-intentions and beliefs—that we-intentions are "primitive" because of their distinct phenomenology (*op cit*, p 404)—and undertakes an analysis of his own The critique is persuasive at the intuitive level but, on closer inspection, turns out to be question-begging Searle says that, in cases of collective intentionality, individual I-intentions are "derivative from" we-intentions "in a way we will need to explain" (*op cit*, p 403) He then analyzes collective intentions with reference to a case in which Jones and Smith are preparing a hollandaise sauce together, Jones stirring while Smith pours Searle suggests that the we-intention to make the sauce by means of Jones's stirring is like an intention to fire a gun by means of pulling the trigger The distinctive problem of collective intentions is how to characterize each individual's intention without having *either* Jones's individual intention causing the whole joint action *or* having Jones intend Smith's action, and vice versa On Searle's analysis, Jones's description of what is going on is "We make the sauce by means of Me stirring and You pouring," and the intention in Jones's mind is "We intend to make the sauce by means of Me stirring" (*op cit*, p 412) The idea seems to be that the I-intention to stir is *part of* the we-intention to make the sauce This is not quite the *derivation* of I-intentions from we-intentions that Searle said we needed

Whatever one makes of this analysis, it does not resolve the problem with which Searle began We can still ask why, in the Hawk-Dove example, P1 and P2 do not have a collective intention with respect to (*hawk, dove*) What is wrong with saying that, in P1's mind, there is a we-intention to play the combination (*hawk, dove*) by means of P1 playing *hawk* and P2 playing *dove*? Searle asserts that "the notion of a we-intention implies the notion of *cooperation*" (*op cit*, p 406) and says that cooperation is construed in terms of "collective goals" (*op cit*, pp 405, 411), so he might answer that, in playing the strategy combination (*hawk, dove*), P1 and P2 are not "cooperating" in pursuit of a "collective goal" But those concepts are left unanalyzed

So Searle's analysis, like Tuomela and Miller's, distinguishes between genuinely collective intentions and the individual intentions of Nash equilibrium only by appealing to an intuitive understanding of the concepts of "cooperation" (or "joint action") and "collective goal" or ("joint goal") From the perspective of the literature of collective intentionality, the problem is to find a way of making the cooperative pursuit of a collective goal a property of the corresponding intention The difficulties that these writers have found in grappling with this problem suggest that a different line of approach might be more useful Our approach will be to analyze the nature of practical reasoning associated with the pursuit of a collective goal, and then to investigate the intentions that it produces

Bratman offers a rather different account of collective intentionality in his analysis of "shared cooperative activity" For Bratman, the key feature of shared cooperative activity is that, for each agent, I must intend that we *J* "in part *because of* your intention that we *J*"[5] This must be known by both agents (In "Shared Intention," there is an explicit common knowledge condition This much gets us the core analysis of View 3 in that paper) It is not clear that Bratman's basic analysis, any more than Tuomela and Miller's or Searle's, excludes Nash equilibrium play in Hawk-Dove In Nash equilibrium, each individual's action is adapted to her beliefs about the actions of the other and, since these are intentional actions, that is equivalent to saying that each individual's intention is adapted to her beliefs about the intentions of the other If P1 intends the Nash equilibrium outcome

---

[5] "Shared Cooperative Activity," p 333 (with italics), "Shared Intention," p 104 (without italics)

(*hawk, dove*), and believes that this outcome will come about in part because P2 will play dove, and that P2 will play dove because he intends to, and that P2 intends to play dove because he intends the outcome (*hawk, dove*) and because he believes that (*hawk, dove*) will come about in part because P1 will play *hawk*, and so on, then it seems that playing (*hawk, dove*) will be a shared cooperative activity in Bratman's sense. Bratman might object that it is too glib to interpret P1's intending that *J* come about "because of" P2's intention as the idea that P1 believes that P2 has the corresponding intention and acts on the basis of this belief. But his expansion of it—"I intend that our performance of the joint activity be in part explained by your intention that we perform the joint activity, I intend that you participate as an intentional agent in a joint activity that, as I know, you too intend"[6]—is opaque. The subsequent discussion reveals that P1 must intend the "efficacy" of P2's intention, or that the outcome is achieved through P2's intention to achieve it. It is not clear how this would exclude there being a collective intention with respect to the pair of strategies (*hawk, dove*).

An alternative response is that the Hawk-Dove case is excluded from the scope of Bratman's analysis because the two players move once only, and simultaneously. Bratman uses what he calls a "planning conception of intention,"[7] whereby an intention is an action-guiding mental state that is maintained over an interval of time, the cooperative activities that he has in mind are ones in which individuals coordinate their actions over time. Thus, for example, someone might have an individual intention to paint her house, this intention would then guide the formation of "sub-plans" for buying paint, cleaning walls, and so on. Analogously, Bratman argues that "we" might have a collective intention to paint the house which guides the formation of "my" and "your" sub-plans.[8] In game-theoretic terms, this transforms the problem into one of dynamic choice. We might represent the shared cooperative activity as a possible outcome of an extended game comprising a sequence of "stage games" (for instance, a series of Hawk-Dove games), with a stage-game strategy counting as a "subplan." Bratman says that the subplans of players A and B for a joint action *J mesh* if there is some way that "we could *J* that would involve the successful execution of those subplans."[9] So the intentions behind a sequence of Nash equilibria in the stage games (for instance, "we intend to perform the sequence in which, in every stage game, A plays *hawk*, B plays *dove*") would be a collective intention

---

[6] "Shared Intention," p. 104.

[7] "Shared Cooperative Activity," pp. 330–31.

[8] Adding these subplans gets the schema on pp. 333–34 of "Shared Cooperative Activity," and View 4 in "Shared Intention."

[9] "Shared Cooperative Activity," p. 332, "Shared Intention," p. 106.

However, it is clear that Bratman would not be happy with this characterization of his views He says that his view implies more than that each player "sees the other's intentions as data for [her] deliberations, albeit as data that are potentially affected by [her] own decision"—in other words, it requires more than Nash equilibrium It requires that each agent "aims at the efficacy of the intention of the other" and "embrace[s] as her own end the efficacy of the other's relevant intention "[10] How he would discriminate between shared intentions and Nash equilibrium behavior is clearer in his "Shared Cooperative Activity" paper, which includes further conditions, particularly that each agent's intention is "minimally cooperatively stable" (*op cit*, p 338) This requires that there are at least some circumstances in which she would help the other agent if a new problem arose Bratman also specifies that the agents' attitudes have to be appropriately connected to the joint action which, he says, explicitly distinguishes shared cooperative activity from the kind of "pre-packaged cooperation" that can be represented by the choice of strategies in a game If our roles in an activity are fully determined in advance and we can each carry out our part with no further interaction then, whilst the planning stage may have been a shared cooperative activity, the noninteractive performance of the plan itself is pre-packaged cooperation Although both shared cooperative activity and pre-packaged cooperation involve our bringing about that we *J* through "mutual responsiveness of intention," shared cooperative activity involves the additional feature of "mutual responsiveness in action " To illustrate mutual responsiveness in action, Bratman gives the example of singing a duet, each singer accommodating herself to the actions of the other The intuition is clear enough, but there is no analysis of how mutual responsiveness works So the status of mutual responsiveness is similar to that of "collective intention" and "joint action" prior to Bratman's work—something we have an intuitive understanding of, but want an analysis of

We are also left wondering how Bratman's project relates to other analyses of collective intentions He differentiates it from the analyses provided by Tuomela and Miller and Searle by saying that his shared intentions are states of affairs made up of the interrelated intentions and beliefs of the people who share them, they are not intentions of a special kind, held by individual agents [11] However, we see Bratman's use of the planning conception of intention as a much more fundamental distinction, in that it makes collective intentions prior to practical reasoning about cooperative activity We will argue that his theory may be best understood as one which analyzes group agency in terms of the mental states of the individual agents that compose the group, and which provides the background circumstances in which individual agents use team reasoning

---

[10] "Shared Intention," p 107

II TEAM REASONING

In decision theory, it is almost universally presupposed that agency is invested in individuals each person acts on her own preferences and beliefs A person's preferences may take account of the effects of her actions on other people, she may, for example, be altruistic or have an aversion to inequality Still, these are *her* preferences, and she chooses what *she* most prefers Opposing this orthodoxy is a small body of literature which allows groups or "teams" of individuals to count as agents, and which seeks to identify distinctive modes of team reasoning that are used by individuals as members of teams This idea has been around for some time, having been proposed in different forms by David Hodgson,[12] Donald Regan,[13] Margaret Gilbert,[14] Susan Hurley,[15] Robert Sugden,[16] Martin Hollis,[17] Michael Bacharach,[18] and Elizabeth Anderson [19] One motivation for theories of team reasoning is that there are games that are puzzles for orthodox decision theory, in the sense that there exists some strategy that is at least arguably rational and that a substantial number of people play in real life, but whose rationality decision theory cannot explain and whose play it cannot predict These are puzzles of *cooperation*, in that the strategies offer what Bacharach calls *scope for common gain*, or the possibility of making a Pareto improvement (that is, at least one player is made better off and no player is made worse off) on an outcome predicted by a standard solution concept (*op cit*)

---

[11] "Shared Intention," pp 99, 103

[12] Hodgson, *Consequences of Utilitarianism* (New York Oxford, 1967)

[13] Regan, *Utilitarianism and Cooperation* (New York Oxford, 1980)

[14] Gilbert, *On Social Facts* (New York Routledge, 1989)

[15] Hurley, *Natural Reasons* (New York Oxford, 1989)

[16] Sugden, "Thinking as a Team Toward an Explanation of Nonselfish Behavior," *Social Philosophy and Policy*, x (1993) 69–89, "The Logic of Team Reasoning," *Philosophical Explorations*, vi (2003) 165–81

[17] Hollis, *Trust within Reason* (New York Cambridge, 1998)

[18] Bacharach, "Interactive Team Reasoning A Contribution to the Theory of Cooperation," *Research in Economics*, LIII (1999) 117–47, *Beyond Individual Choice*, Natalie Gold and Robert Sugden, eds , (Princeton University Press, 2006)

[19] Anderson, "Unstrapping the Straitjacket of 'Preference' A Comment on Amartya Sen's Contributions to Philosophy and Economics," *Economics and Philosophy*, xvii (2001) 21–38

One such puzzle is the Prisoner's Dilemma Conventional game theory predicts that players will always choose *defect*, while in fact many players choose *cooperate* There is scope for common gain because if both players choose *cooperate* then they are both better off than if they both choose *defect* In experiments in which people play the Prisoner's Dilemma for money, anonymously and without repetition, the proportion of participants choosing *cooperate* is typically between 40 and 50 percent[20] The theory is failing to explain observed behavior in games There is a parallel problem for normative game theory The theory prescribes *defect*, but many people have the strong intuition that *cooperate* is the rational choice Of course, it is open to the game theorist to argue that that intuition is mistaken, and to insist on the normative validity of the standard analysis In doing so, the game theorist can point out that any individual player of the Prisoner's Dilemma does better by choosing *defect* than by choosing *cooperate*, irrespective of the behavior of her opponent In other words, each individual player can reason to the conclusion "The action that gives the best result *for me* is *defect*" But, against that, it can be said with equal truth that the two players of the game both do better by their both choosing *cooperate* than by their both choosing *defect* Thus, each player can also reason to the conclusion "The pair of actions that gives the best result *for us* is not (*defect*, *defect*)"[21] It seems that normative argument between these two positions leads to a stand-off

A second puzzle is the game of Hi-Lo This is a game in which each player chooses one element from the same set of *labels* As in a pure coordination game, the two players get the same strictly positive payoff if both choose the same label and (0, 0) otherwise However (and in contrast to a pure coordination game, in which all labels give the same payoff), there is one label that gives a strictly higher payoff from coordination for each player than the others Figure 3 shows a simple version of Hi-Lo, in which there are just two labels, *high* and *low*

Hi-Lo combines features of pure coordination games and the Prisoner's Dilemma Like a pure coordination game, this is a *common interest game*—that is, a game in which the interests of the players are perfectly

---

[20] David Sally, "Conversation and Cooperation in Social Dilemmas A Meta-Analysis of Experiments from 1958 to 1992," *Rationality and Society*, VII (1995) 58–92

[21] In order to conclude that (*cooperate*, *cooperate*) is the best pair of strategies for them, the players have to judge the payoff combinations (4, 1) and (1, 4) to be worse "for them" than (3, 3)

		Player 2	
		high	***low***
Player 1	***high***	**2, 2**	**0, 0**
	low	**0, 0**	**1, 1**

Figure 3 Hi-Lo

aligned, signalled by the fact that, in each cell of the payoff matrix, the two players' payoffs are equal to one another There are two pure-strategy Nash equilibria, each associated with a different label and coming about if both players choose that label In this sense, Hi-Lo poses a coordination problem each player wants it to be the case that they both choose the same label The crucial difference from a pure coordination game is that, in Hi-Lo, one of the equilibria is strictly better than the other for both players At first sight, this makes the coordination problem in Hi-Lo trivial it seems obvious that the players should coordinate on the equilibrium they both prefer, namely (*high, high*)

Hi-Lo shares with the Prisoner's Dilemma the feature that, of the outcomes that occur if both players choose the same label, one is better than the other for both players In this sense, Hi-Lo poses a cooperation problem there is scope for common gain because both players benefit by their both choosing *high* rather than *low* just as, in the Prisoner's Dilemma, both players benefit by their both choosing *cooperate* rather than *defect* The difference is that in the Prisoner's Dilemma, (*cooperate, cooperate*) is not a Nash equilbrium while in Hi-Lo, (*high, high*) is It might seem that, because of this difference, the cooperation problem in Hi-Lo is trivial too

Certainly, Hi-Lo does not pose practical problems for ordinary people, either individually or collectively In experiments in which participants play Hi-Lo games, and in which the *high* and *low* strategies are given neutral labels, the overwhelming majority choose *high*[22] But Hi-Lo presents a fundamental problem for game theory From the assumptions that the players are perfectly rational (in the normal sense of maximizing expected payoff) and that they have common knowledge of their rationality, we cannot deduce that each will choose *high* Or, expressing the same idea in normative terms, there is no sequence of steps of valid reasoning by which perfectly rational players can arrive at the conclusion that they ought to choose *high* Many people find this claim incredible, but it is true It is true because, from the assumption of rationality, all we can infer is that each player chooses the strategy that maximizes her expected payoff, given her beliefs about what the other player will do All we can say in favor of *high* is that, if either player expects the other to choose *high*, then it is rational for the first player to choose *high* too, thus, a shared expectation of *high*-choosing is self-fulfilling among rational players But exactly the same can be said about *low* Intuitively, it seems obvious that each player should choose *high* because both prefer the outcome of (*high*, *high*) to that of (*low*, *low*), but that 'because' has no standing in the formal theory[23]

---

[22] Nicholas Bardsley, Judith Mehta, Chris Starmer, and Sugden, "The Nature of Salience Revisited Cognitive Hierarchy Theory *versus* Team Reasoning" (unpublished manuscript, 2006), report an experiment where 56 student subjects were presented with two Hi-Lo games In one game, the ratio of the money payoffs to *high* and *low* was 10 1, in the other it was 10 9 In each case, 54 subjects (96 percent) chose *high*.

If we are prepared to relax the classical assumption of perfect rationality, it is not particularly difficult to construct theories which purport to explain the choice of *high* After we have stripped out any information contained in their labels, the only difference between the *high* and *low* strategies is that *high* is associated with higher payoffs, because of this, most plausible theories of imperfect rationality predict that *high* is more likely to be chosen than *low*[24] But it seems unsatisfactory to have to invoke assumptions about imperfections of rationality in order to explain behavior in such a transparently simple game as Hi-Lo If we find that standard game-theoretic reasoning cannot tell players how to solve the apparently trivial problem of coordination and cooperation posed by Hi-Lo, we may begin to suspect that something is fundamentally wrong with the whole analysis of coordination and cooperation provided by the standard theory Conversely, if we could find a form of reasoning which recommends *high* in Hi-Lo, that might provide the key to solving the problem posed by the Prisoner's Dilemma

The source of both puzzles seems to be located in the mode of reasoning by which, in the standard theory, individuals move from preferences to decisions In the syntax of game theory, each individual must ask separately "What should *I* do?" In Hi-Lo, the game-theoretic answer to this question is indeterminate In the Prisoner's Dilemma, the answer is that *defect* should be chosen Intuitively, however, it seems possible for the players to ask a different question "What should *we* do?" In Hi-Lo, the answer to *this* question is surely "Choose (*high, high*) " In the Prisoner's Dilemma, "Choose (*cooperate, cooperate*)" seems to be at least credible as an answer Theories of team agency try to reformulate game theory in such a way that "What should we do?" is a meaningful question The basic idea is that, when an individual reasons as a member of a team, she considers which *combination* of actions by members of the team would best promote the team's objective, and then performs her part of that combination The rationality of each individual's action derives from the rationality of the joint action of the team In the same way that individual reasoning leads to individual intentions, it is natural to suppose that team reasoning leads to collective intentions In this section of the paper we show how the we-intentions of individual agents would arise as the result of team reasoning

---

[23] For fuller statements of this argument, see Hodgson, *Consequences of Utilitarianism*, Sugden, "Thinking as a Team Toward an Explanation of Nonselfish Behavior", or Bacharach, *Beyond Individual Choice*.

[24] For example, suppose that each player believes that his opponent is just as likely to choose one strategy as the other Then both will choose *high* Or suppose that each player believes that his opponent believes that he is just as likely to choose one strategy as the other Then each player will expect his opponent to choose *high*, and so choose *high* as a best reply Or

We represent team reasoning explicitly, as a *mode of reasoning* in which propositions are manipulated according to well-defined rules—an approach that has previously been used by Natalie Gold and Christian List[25] Our basic building block is the concept of a *schema of practical reasoning*, in which conclusions about what actions should be taken are inferred from explicit premises about the decision environment and about what agents are seeking to achieve In propositional logic, a rule of inference—a rule that allows us to derive conclusions from premises—is valid if, whenever the premises are true, so are the conclusions that are derived from them One can formulate principles of practical reasoning which satisfy analogous criteria of validity Following Bacharach, we define a mode of practical reasoning for a given class of games as *valid* if it is *success-promoting* given any game in that class, it yields only choices which tend to produce success, as measured by game payoffs[26] The fundamental idea is that practical reasoning infers conclusions about what an agent ought to do from premises which include propositions about what the agent is seeking to achieve Such reasoning is *instrumental* in that it takes the standard of success as given, its conclusions are propositions about what the agent should do in order to be as successful as possible according to that standard If the agent is an individual person, the reasoning is *individually instrumental*. Here is a simple example of individually instrumental reasoning, in which (1) to (4) are premises and the proposition below the line is the conclusion.

---

[25] Gold and List, "Framing as Path-Dependence," *Economics and Philosophy*, xx (2004) 253–77

[26] Bacharach, *Beyond Individual Choice*, pp 7–10

Schema 1. Individual Rationality

(1) I must choose either *left* or *right*.
(2) If I choose *left*, the outcome will be O_1
(3) If I choose *right*, the outcome will be O_2
(4) I want to achieve O_1 more than I want to achieve O_2

I should choose *left*

In our analysis, we will interpret the "official" payoffs of a game (as represented in the matrix which defines the game, Figures 1, 2 and 3 being examples) as specifying what the players want to achieve as individuals or, equivalently, what counts as success for them. Following the conventions of game theory, we will treat payoffs as utility indices in the sense of expected utility theory so that, in situations of uncertainty, a player's success is measured by the expected value of her payoff.

Now consider the following schema, in which (*left*, *right*) denotes the pair of actions "I choose *left*, you choose *right*".

Schema 2. Collective Rationality

(1) We must choose one of (*left*, *left*), (*left*, *right*), (*right*, *left*) or (*right*, *right*)
(2) If we choose (*left*, *left*) the outcome will be O_1
(3) If we choose (*left*, *right*) the outcome will be O_2
(4) If we choose (*right*, *left*) the outcome will be O_3
(5) If we choose (*right*, *right*) the outcome will be O_4
(6) We want to achieve O_1 more than we to achieve O_2, O_3, or O_4

We should choose (*left*, *left*)

Is this schema valid? Given the symmetries between Schemata 1 and 2, it seems that, if one is valid, so too is the other. Both present instrumental practical reasoning, where an agent deliberates about what to do in order to achieve its goals. The only difference is the level of agency. If Schema 1 represents valid reasoning for an individual agent, Schema 2 is its parallel for a group agent.

As far as we can see, the only grounds for objecting to Schema 2 while accepting Schema 1 is to claim that the concept of group agency itself is incoherent. This claim can construed in two different ways.

On the first construal, the claim is that expressions such as 'we want to achieve' and 'we should choose' are mysterious, because wanting and choosing are mental states and mental states must be located in individual minds. In the rest of this section of the paper, we try to answer this objection by showing how the "we" concepts of Schema 2 correspond with "I" concepts. We assert that Schemata 1 and 2 are both forms of valid instrumental reasoning and that they are valid *for agents*, defined as those entities that use these modes of reasoning (individual agents in one case, group agents in the other).

On the second construal, it is a claim about social ontology. The counter-claim that groups *can* be agents in this metaphysical sense is defended by Carol Rovane[27] and Philip Pettit.[28] For both Rovane and Pettit, an agent is characterized by a commitment to "rational unity", it can be any entity that has the ability to form states that play the role of intentional attitudes, such as judgments and beliefs, and that can take steps to ensure that these states are consistent with each other. If the defining feature of agency is the ability to engage in rational deliberation, then the bounds of agency need not coincide with the individual. Rovane and Pettit's is a normative concept of agency, it presupposes principles of rationality which are constitutive of being an agent, and derives the metaphysical possibility of group agency from these principles.

In this paper, we do not make any ontological or metaphysical claims, rather, we describe a type of reasoning which the members of group agents use. This means that we are using a more inclusive definition of group agent than Rovane and Pettit. On their definition some of the examples of ephemeral groups in the collective intentions literature, such as two strangers pushing a car, will not exhibit agency. But these groups may use team reasoning. In section III, we will present a range of alternative accounts of what group agency is and how it comes about, some of these accounts are compatible with Rovane and Pettit's definition, while others are not. In this section, we confine ourselves to showing that there are practical reasoning schemata that use "we" concepts, which individual people can (and, we claim, sometimes do) use, which are similar in structure to modes of individual practical reasoning.

We begin by noting what might seem to be a potential contradiction between Schemata 1 and 2. For example, consider the

---

[27] Rovane, *The Bounds of Agency* (Princeton University Press, 1988).

[28] Pettit, *Weakness of the Will and Practical Irrationality* (New York: Oxford, 2003).

Prisoner's Dilemma. Using a variant of Schema 1, each player can apparently reason to "I should choose *defect*", but using a variant of Schema 2, each can apparently reason to "We should not choose (*defect, defect*)." We suggest that the resolution of this problem is that, properly understood, the two sets of premises are mutually inconsistent.[29] Thus, while both schemata are valid, they cannot be used simultaneously to infer contradictory conclusions. The premises of Schema 1 presuppose that *I* am an agent, pursuing *my* objectives. Those of Schema 2 presuppose that *we* make up a single unit of agency, pursuing *our* objectives. But instrumental practical reasoning presupposes a unit of agency. If I am to reason instrumentally, I cannot simultaneously think of myself both as a unit of agency in my own right and as part of a unit of agency which includes you.[30]

We can make this feature of practical reasoning more transparent by writing schemata in forms which include premises about agency. Consider any situation in which each of a set S of individuals has a set of alternative *actions*, from which he must choose one. A *profile* of actions assigns to each member of S one element of his set of alternative actions. For each profile, there is an *outcome*, understood simply as the state of affairs that comes about (for everyone) if those actions are chosen.[31] We define a *payoff function* as a function which assigns a numerical value to every outcome. A payoff function is to be interpreted as representing what some specific agent wants to achieve: if one outcome has a higher numerical value than another, then the relevant agent wants to achieve the first more than he (or she, or it) wants to achieve the second.[32] Now consider any individual *i*, and any set of individuals G, such that i is a member of G and G is a weak subset of S We will say that i *identifies with* G if i conceives of G as a unit of agency, acting as a single entity in pursuit of some single objective Finally, we define *common knowledge* in the usual way a proposition x is common knowledge in a set of individuals G if (i) x is true, (ii) for all individuals i in G, i knows x, (iii) for all individuals i and j in G, i knows that j knows x, (iv) for all individuals i, j, and k in G, i knows that j knows that k knows that x, and so on

---

[29] Gold shows a technical sense in which this is so, within a model of reasoning involving the manipulation of propositions in *Framing and Decision Making: A Reason-Based Approach* (Unpublished D.Phil. thesis, University of Oxford, 2005).

[30] This is not to deny the psychological possibility that a person might simultaneously experience motivational or affective pulls towards both individual and group identity. Our claim is merely that such conflicting pulls *cannot be resolved by instrumental reasoning*. Consider an analogous case in conventional choice theory. What if an individual faces a choice between two options, feels motivational pulls towards each of them, but cannot settle on any firm preference (or on a firm attitude of indifference)? Clearly, this case is psychologically possible, but if a person is unsure of her own objectives, instrumental rationality cannot tell her what they should be.

[31] In game-theoretic language, this is a *game form*. A game form consists of a set of players, a set of alternative strategies for each player, and, for each profile of strategies that the players might choose, an outcome. In contrast, a *game* is normally defined so that, for each profile of strategies, there is a vector of numerical payoffs, one payoff for each player.

[32] Thus, the "official" payoffs of a game are the values of payoff functions which represent what each player would want to achieve as an individual agent.

Letting A stand for any profile and U for any payoff function, consider the following schema

Schema 3 Simple Team Reasoning (from a group viewpoint)

(1) We are the members of S
(2) Each of us identifies with S
(3) Each of us wants the value of U to be maximized
(4) A uniquely maximizes U

Each of us should choose her component of A

This schema captures the most basic features of team reasoning Notice that, because of (2), the schema does not yield any conclusions unless all the members of S identify with this group Because of (4), the schema yields conclusions only when a profile that is the unique maximizer of the team payoff function exists We will not address the question of what a team reasoner should do when this is not the case but, for our purposes, the answer is not essential Notice also that we can apply Schema 3 in cases in which S contains only one individual In this case, S can be written as {myself} (1) then becomes "I am the only member of the set {myself} " (2) reduces to "I identify with {myself}," which amounts to saying that the reasoning individual views herself as an agent And then the schema represents straightforward practical reasoning by an individual agent. Thus, Schema 3 encompasses both individual and team reasoning

Schema 3 represents a mode of reasoning that can be used by people *as a group* What does it mean for a number of people to reason as a group? One way to make sense of this is to imagine those people in an open meeting, at which each of a set of premises is announced, and acknowledged as true by each person Then, the inference to be drawn from those premises is announced, and acknowledged as valid by each person In such a setting, it is common knowledge among the members of the group that each of them accepts the relevant premises That this is common knowledge does not need to be stated explicitly in the schema, it is not an additional premise, but a presupposition of the whole idea of reasoning as a group

For many purposes, however, it is more convenient to represent team reasoning from the viewpoint of an individual team member If we adopt this approach, Schema 3 can be rewritten as follows

Schema 4 Simple Team Reasoning (from an individual viewpoint)

(1) I am a member of S
(2) It is common knowledge in S that each member of S identifies with S
(3) It is common knowledge in S that each member of S wants the value of U to be maximized
(4) It is common knowledge in S that A uniquely maximizes U

I should choose my component of A

Schema 4 represents team reasoning as a mode of reasoning that can be used by an individual group member If it is common knowledge that every member of the group identifies with the group and wants the group's payoff function to be maximized, and if it is also common knowledge that a particular profile of actions A uniquely maximizes that function, then premises (1) to (4) are accessible to each member Each can then reason independently to the conclusion that she should choose her component of A

Team reasoning was originally introduced to explain how, when individuals are pursuing collective goals, it can be rational to choose strategies that realize scope for common gain But it also provides an account of the formation of collective intentions Team reasoning, as represented by Schema 4, results in the formation of intentions An individual who accepts premises (1) to (4) and infers "I should choose my component of A" has reason to form the intention to choose that component That intention, if formed, is a mental state of an individual person Nevertheless, references to the group are noneliminable parts of the reasoning process that led to the formation of the intention Thus, it is natural to regard the intentions that result from team reasoning as collective intentions

Interpreted in this way, our analysis does everything Searle asked for The "we-ness" of we-intentions is primitive, in the sense that group agency is a noneliminable part of the reasoning process by which these intentions are formed The presence of group agency also explains the distinctive phenomenology of we-intentions Collective goals (in the form of the group payoff function U) play a fundamental role in the analysis I-intentions are derived from we-intentions in the move from Schema 2, through Schema 3, to Schema 4 The resulting individual intention dovetails with an alternative to Searle's analysis proposed by Nicholas Bardsley,[33] which is intended to be compatible with team reasoning.[34]

Our analysis allows us to distinguish collective intentions from the individual intentions that exist in Nash equilibrium, and thus to solve the general problem we identified in the literature of collective intentions. This distinction can be made by referring to the unit of agency in the reasoning process that led to the formation of the relevant intentions. Nash equilibrium, as normally understood, is a relationship between the strategy choices of players who reason as individuals (or, equivalently, each of whom identifies with the one-person group which contains only himself). Correspondingly, the intentions which underlie Nash equilibrium are the result of individual reasoning. In contrast, collective intentions are the product of team reasoning.

As an illustration, we return to the Hawk-Dove example discussed in section I. In this example, it is common knowledge between P1 and P2 that, in Hawk-Dove interactions, the player in the position of P1 almost always chooses *hawk* and the one in the position of P2 almost always chooses *dove*. Expecting P2 to play *dove*, P1 forms the intention to play *hawk*. Expecting P1 to play *hawk*, P2 forms the intention to play *dove*. We posed the question: Does each player have a we-intention with respect to (*hawk*, *dove*)? To answer this question, we investigate the reasoning process by which the intentions were formed. Given his belief that P2 will almost certainly choose *dove*, P1 reasons as follows:

(1) I must choose either *hawk* or *dove*.
(2) If I choose *hawk*, the outcome will (almost certainly) be a payoff of 3.
(3) If I choose *dove*, the outcome will (almost certainly) be a payoff of 2.
(4) I want to achieve a payoff of 3 more than I want to achieve a payoff of 2.

I should choose *hawk*.

Reasoning similarly, P2 reaches the conclusion "I should choose *dove*." Both chains of reasoning follow the "individual rationality" model of Schema 1 We conclude that P1's and P2's intentions with respect to (*hawk, dove*) are not collective

---

[33] Bardsley, "On Collective Intentions: Collective Action in Economics and Philosophy," *Synthese* (forthcoming, 2007).

[34] Bardsley's alternative to "We intend to make the sauce by means of me stirring" as Jones's intention would be "I intend my part of the combination (Jones stirs, Smith pours) in circumstances that you and I have this very intention, all of which is to make the sauce." In Bardsley's analysis, Smith's intention has ***exactly the same*** sense as Jones's (although 'you' and 'I' refer to a different people in the two cases).

It is an implication of our analysis that a given pattern of behavior can be intended either individually or collectively, depending on the reasoning which led to it For example, compare the following two stories about Hi-Lo In the first story, P1 and P2 reason as individuals Each forms the belief that the other is just as likely to play one strategy as the other Since this implies that the expected individual payoff from choosing *high* is greater than that from choosing *low*, each forms the intention to choose *high* In this case, (*high, high*) is a Nash equilibrium in the standard sense, and the players' intentions are not collective In the second story, each of P1 and P2 identifies with the group {P1, P2} and takes the group payoff function to be one which is uniquely maximized by the profile (*high, high*), all of this is common knowledge between them Each forms the intention to play *high* as his component in this profile These intentions are collective[35]

III THEORIES OF GROUP AGENCY

The pure theory of team reasoning, as presented in the previous section of this paper, presupposes that there can be group agency, but it is not reliant on any particular theory of how group agency comes about or of what the group agent should take as its goals The literature of team reasoning offers a range of possible answers to these questions In this section, we briefly review various theories of group agency Our aim is not to adjudicate between them, but merely to indicate the variety of philosophically significant ways in which group agency can be interpreted In the process, we locate Bratman's analysis within our framework

III 1 Team Agency Required by Rationality or Morality The first theorists to discuss team reasoning did so in the context of moral and rational requirements on action In these accounts, what we have called "identifying with" a group is construed as an individual's response to such requirements

Hodgson was the first person to use the Hi-Lo game, as part of an argument that rule utilitarianism does not reduce to act utilitarianism (*op cit*) Expanding on this argument, Regan proposed a form of team reasoning in his theory of *cooperative utilitarianism* Regan's theory is normative, it is commended to all of us in our capacities as rational and moral agents The fundamental principle of this theory is that "what each agent ought to do is to co-operate, with whoever else is co-operating, in the production of the best consequences possible given the behaviour of non-co-operators" (*op cit*, p 124) In a world in which everyone is a cooperative utilitarian, Regan's rational and moral agents reason according to Schema 4, S is the set of all agents, and the value of U is a measure of the overall good of the world

---

[35] Because conventional game theory does not recognize group agency, we cannot appeal to convention to resolve the question of whether (*high, high*) is properly called a "Nash equilibrium" in this case It is a Nash equilibrium in relation to the official payoffs of the game, but the players are not motivated by these payoffs in the standard way

Regan's theory also gives recommendations for cases in which not everyone is a cooperative utilitarian The logic of these recommendations can be represented by a variant of team reasoning called *restricted team reasoning* by Bacharach [36] As in Schema 4, S denotes the group with which team reasoners identify However, it is allowed that not all members of S in fact identify with S. The set of those members of S (the "team" for the purposes of team reasoning) who do so identify is denoted by T Let A_T be a profile of actions for the members of T Then restricted team reasoning is represented by the following schema

Schema 5 Restricted Team Reasoning

(1) I am a member of T
(2) It is common knowledge in T that each member of T identifies with S
(3) It is common knowledge in T that each member of T wants the value of U to be maximized
(4) It is common knowledge in T that A_T uniquely maximizes U, given the actions of nonmembers of T

I should choose my component of A_T

In cooperative utilitarianism, each of us is told to join with as many others as are willing to do the same, and to cooperate with them in trying to achieve the overall good of the world In terms of Schema 5, S is the set of all people, T is the set of cooperative utilitarians, and the value of U is a measure of overall goodness

The idea that team reasoning can be required by morality is not limited to utilitarians Anderson presents a Kantian argument for the rationality of team reasoning and for its morality when applied to the universal community of humanity (*op cit*) Anderson presents this idea while explaining how, in games which offer scope for common gain, individuals can rationally treat their joint strategy profile as the unit of selection Her analysis of the "rational basis for committed ac-

[36] Bacharach, *Beyond Individual Choice*

tion" in such games follows the general logic of Schema 4 The analysis applies to cases in which a group of individuals identify themselves as members of that group, and see their actions as jointly advancing a common goal Under these conditions, it is rational for each member of the group to choose her component of the strategy profile that best advances the goal This conclusion is reached by "collective deliberation," structured by the principle that "whatever can count as a reason for action for one member of the collective must count as a reason for all", it is rational by virtue of the individual's having reason "to do my part in what we are willing together" (*op cit*, pp 28–30)

Anderson does not claim that practical reason requires individuals to identify with any particular groups Rather, identity is prior to rational choice, in the sense that "what principle of choice it is rational to act on depends on a prior determination of personal identity, of who one is" (*op cit*, p 30) However, she holds out the hope that there may be "further principles of rational self-identification" which will tell us which groups we should identify with (*op cit*, p 32) She maintains that morality requires us to transcend our various identities and harmonize their demands, by identifying with a community that comprehends them all—the Kantian Kingdom of Ends—but she admits to having no argument that this is rationally required (*op cit*, p 37) Similarly Kantian aspirations can be found in Hollis's sketch of how team reasoning might be used by "citizens of the world" (*op cit*, pp 150–63)

In a similar vein, Hurley proposes that we (as rational and moral agents) should specify agent-neutral goals—that is, goals of which it can simply be said that they ought to be pursued, rather than they ought to be pursued by some particular agent (*op cit*, pp 136–59) Then we should "survey the units of agency that are possible in the circumstances at hand and ask *what the unit of agency, among those possible, should be*", and we should "ask ourselves *how we can contribute to the realization of the best unit possible in the circumstances* " Like Anderson, Hurley does not nominate any particular group as the rational one to identify with, or any particular goal as the rational one to pursue Nevertheless, the idea seems to be that rationality requires each person to choose the unit of agency in which she participates, and that this choice should be governed by goals which are independent of the unit of agency

III 2 Team Agency as the Result of Framing In contrast, Bacharach's theory does not allow the unit of agency to be chosen, and does not admit the concept of a goal that is not the goal of some agent[37] For

---

[37] Bacharach, *Beyond Individual Choice*

Bacharach, whether a particular player identifies with a particular group is a matter of "framing " A *frame* is the set of concepts a player uses when thinking about her situation In order to team reason, a player must have the concept "we" in her frame Bacharach proposes that the "we" frame is induced or *primed* by games which have a property that Bacharach calls *strong interdependence* Roughly, a game has this property if it has a Nash equilibrium which is Pareto-dominated by the outcome of some feasible strategy profile [38] Although Bacharach proposes that the perception of this property increases the probability of group identification, he does not claim that games with this property *invariably* prime the "we" frame In particular, although the Prisoner's Dilemma exhibits strong interdependence, Bacharach allows that a player of this game may frame it either as a single-agent problem for "us" or as a game to be played by two separate individual agents

In Bacharach's theoretical framework, this dualism is best represented in terms of *circumspect team reasoning* We now present this mode of reasoning in the form of a reasoning schema As before, let S be the set of individuals with which team-reasoners identify, and let T be any subset of S, interpreted as the set of individuals who in fact identify with S Suppose there is a random process which, independently for each member of S, determines whether or not that individual is a member of T, for each individual, the probability that he is a member of T is ω, where $\omega > 0$ We define a proposition p to be *T-conditional common knowledge* if (i) p is true, (ii) for all individuals i in S, if i is a member of T, then i knows p, (ii) for all individuals i and j in S, if i is a member of T, then i knows that if j is a member of T, then j knows p, and so on (As an illustration imagine an underground political organization which uses a cell structure, so that each member knows the identities of only a few of her fellow-members New members are inducted by taking an oath, which they are told is common to the whole organization Then, if T is the set of members, the content of the oath is T-conditional common knowledge) We define a *protocol* as a profile of actions, one for each member of S, with the interpretation that the protocol is to be followed by those individuals who turn out to be members of T Let P be any protocol The schema is

Schema 6 Circumspect Team Reasoning

(1) I am a member of T
(2) It is T-conditional common knowledge that each member of T identifies with S

---

[38] For a more formal definition, see Bacharach, *Beyond Individual Choice*

(3) It is T-conditional common knowledge that each member of T wants the value of U to be maximized

(4) It is T-conditional common knowledge that P uniquely maximizes U, given the actions of nonmembers of T

I should choose my component of P

Bacharach applies this theory to the Prisoner's Dilemma, setting S = {P1, P2} and interpreting U as a measure of the value of the outcome of the game to P1 and P2 together, he assumes that U is maximized when both players choose *cooperate* If the value of U is higher when just one player chooses *cooperate* than when neither does, or if the value of ω is sufficiently high, the uniquely optimal protocol is (*cooperate, cooperate*) This gives a model in which players of the Prisoner's Dilemma choose *cooperate* if the "we" frame comes to mind, and *defect* otherwise Bacharach offers this result as an explanation of the observation that, in one-shot Prisoner's Dilemmas played under experimental conditions, each of *cooperate* and *defect* is usually chosen by a substantial proportion of players

Bacharach claims that Schema 6 is valid, with the implication that, for any given individual, *if she identifies with S and wants U to be maximized*, it is instrumentally rational for her to act as a member of T, the team of like-minded individuals He does not claim that she *ought* to identify with any particular S, or that she *ought* to want any particular U to be maximized In the theory of circumspect team reasoning, the parameter ω is interpreted as a property of a psychological mechanism—the probability that a person who confronts the relevant stimulus will respond by framing the situation as a problem "for us " The idea is that, in coming to frame the situation as a problem "for us," an individual also gains some sense of how likely it is that another individual would frame it in the same way, in this way, the value of ω becomes common knowledge among those who use this frame

III 3 Team Agency and Assurance Up to now, we have treated schemata of practical reasoning as modes of *valid* reasoning, 'validity' being interpreted instrumentally Thus, team reasoning has been presented as an element of the theory of *rational* choice Sugden offers an alternative approach, presenting a "logic of team reasoning" without making any claims for its validity[39] In this analysis, a "logic" is merely an internally consistent system of axioms and inference rules

---

[39] Sugden, "The Logic of Team Reasoning "

An individual actor may *endorse* a particular logic, thereby accepting as true any conclusions that can be derived within it, but the theorist need not take any position about whether the axioms of that logic are "really" true or whether its inference rules are "really" valid Team reasoning is then represented as a particular inference rule which, as a matter of empirical fact, many people endorse

Sugden is particularly concerned with the following *assurance* question When a team reasoner chooses her component of the profile that maximizes her team's payoff, does she have the assurance that other members of the team are choosing their components too? If one asserts (as Sugden does not) that the relevant schema of team reasoning is valid, one can give the following straightforward answer if the team members know one another to be rational, this knowledge provides the necessary assurance To see why, consider the simplest schema of team reasoning, Schema 4 Notice that this schema tells an individual member of S to choose his component of the U-maximizing profile only in situations in which it also tells the other members to choose theirs Thus, if each member is rational in the sense of being capable of valid practical reasoning and having the motivation to act on the conclusions it generates, each will act on the conclusions of Schema 4, as applied to his case And, since it is common knowledge in S that everyone identifies with S, each player has the resources to work all this out So, whenever Schema 4 tells an individual to choose his component of the U-maximizing profile, that individual has the assurance that the others, *if rational*, will choose theirs too [40] But since Sugden's approach does not acknowledge agent-neutral concepts of "validity" and "rationality," assurance has to be generated in a different way

Following David Lewis[41] and Robin Cubitt and Sugden,[42] Sugden uses a theoretical framework in which the central concept is *reason to believe* To say that a person has reason to believe a proposition *p* is to say that *p* can be inferred from propositions than she accepts as true, using rules of inference that she accepts as valid On the analogue of the definition of common knowledge, there is *common reason to believe* a proposition p in a set of individuals T if (i) for all individuals i in T, i has reason to believe p, (ii) for all individuals i and j in T, i has reason to believe that j has reason to believe p, (iii) for all individuals i, j, and k in T, i has reason to believe that j has reason to believe that k has reason to believe p, and so on.

---

[40] This argument extends to the cases of restricted and circumspect team reasoning, in the sense that each team-reasoning individual has the assurance that other members of T, the "team" of like-minded individuals, will choose their components of the profile or protocol which maximizes the value of the team's payoff function However, a member of T may know with certainty (in the case of restricted team reasoning) or with high probability (in the case of circumspect team reasoning) that other members of S will act as individual agents—for example, by choosing *defect* in the Prisoner's Dilemma

[41] Lewis, *Convention A Philosophical Study* (Cambridge Harvard, 1969)

[42] Cubitt and Sugden, "Common Knowledge, Salience, and Convention," *Economics and Philosophy*, xix (2003) 175–210

The following definition is also useful.[43] Within a set of individuals T, there is *reciprocal reason to believe* that some property q holds for members of T if (i) for all individuals i and j in T, where $i \neq j$, i has reason to believe that q holds for j, (ii) for all individuals i, j, and k in T, where $i \neq j$ and $j \neq k$, i has reason to believe that j has reason to believe that q holds for k, and so on. To see the point of this latter definition, consider the Prisoner's Dilemma and let q be the property 'chooses *cooperate*.' In a schema of practical reasoning which is intended to be used by (say) P1 in deciding how to play the Prisoner's Dilemma, we cannot allow the premise that, in the group {P1, P2}, there is common reason to believe that P1 chooses *cooperate*. That would make it a premise that P1 has reason to believe that he himself will choose *cooperate*, when the whole point of using the schema is to determine which action he should choose. However, we *can* allow the premise that there is reciprocal reason to believe that members of {P1, P2} choose *cooperate*, and there may be circumstances in which such a premise would be natural. For example, suppose that P1 and P2 have played the Prisoner's Dilemma many times before, and on every such occasion, both have chosen *cooperate*. They are about to play again, and there is no obvious difference between this interaction and all its predecessors. Then, by induction, P1 might have reason to believe that P2 will choose *cooperate*. Attributing similar reasoning to his opponent, P1 might have reason to believe that P2 has reason to believe that P1 will choose *cooperate*; and so on.

Sugden's formulation of team reasoning can be represented as the following schema:

Schema 7. Mutually Assured Team Reasoning

(1) I am a member of S.
(2) I identify with S and acknowledge U as its objective.
(3) In S, there is reciprocal reason to believe that every member of S identifies with S and acknowledges U as the objective of S.

---

[43] The concept we define here is similar to that of "reciprocal" belief, used by Frederic Schick, "Surprise, Self-Knowledge, and Commonality," this JOURNAL, XCVII, 8 (August 2000): 440–53.

(4) In S, there is reciprocal reason to believe that every member of S endorses and acts on mutually assured team reasoning.
(5) In S, there is common reason to believe that A uniquely maximizes U.

I should choose my component of A.

This schema is presented merely as a mode of reasoning that any person might (or might not) endorse; a person commits herself to team reasoning by endorsing the schema. Notice that premises (2) and (3) refer to "acknowledging U as the objective of S" rather than "wanting U to be maximized." On Sugden's account, a team reasoner who identifies with a group stands ready to do her part in joint actions in pursuit of the group's objective, but she does not necessarily take this objective as *hers* in the stronger sense of wanting to pursue it even if other members of the group do not reciprocate.

This schema is recursive: premise (4) refers to the endorsement of the schema itself. That this is not circular can be seen from an analogy. Consider an international treaty which includes among its conditions that it will come into effect only if and when it has been ratified by a certain number of nations; once this condition is met, it is binding on every nation that has ratified it. To ratify such a treaty is to make a commitment which is binding from that moment, but which is activated only if enough others make the same commitment. Analogously, to endorse mutually assured team reasoning is to make a unilateral commitment to a certain form of practical reasoning, but this reasoning does not generate any implications for action unless one has assurance that others have made the same commitment. Such assurance could be created by public acts of commitment of a kind we will discuss in the following subsection. But it could also be induced by repeated experience of regularities of behavior in a population. For example, suppose that in some population, some practice of mutual assistance (say, giving directions to strangers when asked) is generally followed in anonymous encounters. Each individual might interpret the existence of the practice as evidence that premises (3), (4) and (5) are true. If so, each individual would be assured that others would choose their components of the U-maximizing profile. But he would still have to decide whether team reasoning was a mode of reasoning that he wanted to endorse.

III.4. Team Agency Produced by Commitment. A final variety of team agency has it that a group is constituted by public acts of promising, or by public expressions of commitment by its members. This latter idea is central to Gilbert's analysis of "plural subjects" (*op. cit.*). Although Gilbert is more concerned with collective attitudes than with collective action, her analysis of how a plural subject is formed might be applied to the formation of teams There are also hints of this approach in the work of Hollis (*op cit*) Hollis suggests that Rousseau's account of the social contract,[44] with its "most remarkable change in man," can be understood as a transition from individual to group agency that takes place through a collective act of commitment

Although Gilbert does not offer an explicit model of collective choice, we suggest that Schema 4 is compatible with her general approach, provided that "identifying with" the group S is understood as some kind of conscious act of commitment On this interpretation, the schema is asserted to be rational, but not in an instrumental sense Rather, the rationality of acting as a member of a team derives from the rationality of fulfilling one's commitments or intentions Focusing on collective attitudes rather than collective actions, Gilbert claims that membership of a plural subject imposes obligations to uphold "our" attitudes This claim is conceptual rather than moral roughly, the idea is that a plural subject is formed by an exchange of commitments, and that to make a commitment is to impose on oneself an obligation to act on it For Gilbert, there is no problem that S (the group with which individuals identify) may be different from T (the set of individuals who in fact identify with S) In commitment-based theories, it is natural to assume that the set of individuals who act as a team is the same as the group with which they identify, provided we can assume that individuals keep their commitments

We suggest that Bratman's analysis of shared intention can be understood as an account of group agency produced by commitment, even though he does not elaborate on how that commitment comes about Given Bratman's "planning concept of intention," we can identify two levels of intentions agents form high-level, *strategic* intentions, which guide the practical reasoning that leads to low-level, *tactical* intentions Strategic intentions set the framework within which subsequent tactical reasoning takes place Bratman's analysis is at the strategic level Recall that, in shared cooperative activity, each agent has the intention that "we" perform some joint activity through the meshing of "my" subplans with "yours " This intention is not linked to any *particular* combination of subplans, rather, it expresses a commitment to engage with the other in a process of "mutual responsiveness" and "mutual support" which is directed towards the meshing of sub-plans *in general* In this sense, the shared intentions are "interlocking" and "end-providing "[45]

---

[44] Rousseau, "The Social Contract" (1762) in Alan Ritter and Julia Conaway Bondanella, eds , *Rousseau's Political Writings* (New York Norton, 1988)

[45] "Shared Cooperative Activity" p 335

In later work, Bratman associates shared intentions with "shared valuing," a shared policy that says what considerations are to be given justifying significance and which provides a background framework for shared deliberation. This would be the analogue of the team payoff function.[46] It seems that, for Bratman, collective intentionality expresses a disposition to reason and act as a member of a group in relation to the objective of executing some broadly-defined joint activity. His analysis leaves open the question of *how*, at the tactical stage, the members of a group coordinate their actions so that together they achieve their joint objective.

Since this is the central question addressed by the theory of team reasoning, the two approaches can be seen as complementary. Bratman's planning conception of shared intention can be thought of as the counterpart in the domain of intentions of group identification in the domain of practical reasons. Seen in this light, Bratman is providing a reductionist account of group agency itself, in terms of the mental states of the individual members of the group.

IV. CONCLUSION

We have identified a general problem in the literature of collective intentions, that of how to distinguish between the "we-intentions" that lie behind cooperative actions and the mutually consistent "I-intentions" that lie behind Nash equilibrium behavior. We have argued that the problem arises because, for this purpose at least, the analyses in the collective intention literature have the wrong starting point. The key difference between the two kinds of intention is not a property of the intentions themselves, but of the modes of reasoning by which they are formed. Thus, an analysis which starts with the intention has already missed what is distinctively collective about it. In our analysis, collective intentions are the product of a distinctive mode of practical reasoning, team reasoning, in which agency is attributed to groups. We have presented the core features of team reasoning in the form of explicit schemata, discussed a range of possible accounts of group agency, and shown how existing theories of collective intentions fit into this framework.

NATALIE GOLD

University of Edinburgh

ROBERT SUGDEN

University of East Anglia

---

[46] Bratman, "Shared Valuing and Frameworks for Practical Reasoning," in R. Jay Wallace, Philip Pettit, Samuel Scheffler, and Michael Smith eds., *Reason and Value. Themes from the Moral Philosophy of Joseph Raz* (New York: Oxford, 2004), pp. 1–27.

Calcutta University Library
ALIPORE
CAMPUS

COLLECTIVITY AND CIRCULARITY*

According to a common claim, a necessary condition for a collective action (as opposed to a mere set of intertwined or parallel actions) to take place is that the notion of collective action figures in the content of each participant's attitudes Insofar as this claim is part of a conceptual analysis, it gives rise to a circularity challenge that has been explicitly addressed by Michael Bratman and Christopher Kutz[1]

I will briefly show how the problem arises within Bratman's and Kutz's analyses, and then proceed to criticize some possible responses, including the ones proposed by Bratman and Kutz My conclusion is that in order to avoid circularity and retain the features that are supposed to make this sort of account attractive, we need a notion of collectivity that does not presuppose intention I suggest that we should make a distinction between collective and noncollective activity merely in terms of dispositions and causal agency There are independent reasons to think that we actually possess such a distinct causal conception of collectivity It is not necessary for the participants in a jointly intentional collective action to possess a stronger notion of their intended collective activity than this In particular, they do not need to possess the concept of a jointly intentional collective action

I THE CIRCULARITY CHALLENGE

Bratman discusses the circularity challenge to his theory of collective action in "Shared Cooperative Activity," and "I Intend That We *J*" (*ibid*) In his account, collective actions spring from shared intentions, and "our shared intention" should be understood in terms of

---

*Versions of this article have been presented at the Higher Seminar in Practical Philosophy in Lund, at ECAP V in Lisbon (2005), and at Collective Intentionality V in Helsinki (2006) I am grateful to participants in these seminars and conferences for valuable criticism and suggestions I am also indebted to Michael Bratman for helpful comments on an earlier version of the paper, to Margaret Gilbert for useful suggestions in conversation, and to anonymous referees for proposals about clarifications This article was prepared as part of a research project for the Swedish Science Council

[1] Bratman, "Shared Cooperative Activity," in *Faces of Intention* (New York Cambridge, 1999), pp 93–109, see p 97 (originally published in *The Philosophical Review*, ci (1992) 327–41), and "I Intend That We *J*," in *Faces of Intention*, pp 142–65, see p 148 (originally published in Raimo Tuomela and Ghita Holmström-Hintikka, eds , *Contemporary Action Theory* (Boston Kluwer, 1997), pp 49–64), Kutz, *Complicity* (New York Cambridge, 2000), see p 86

0022-362X/07/0403/138–56

each of us having an intention of the form 'I intend that we *J*' What is it, then, that I intend when I intend that we *J*? How does this intention of mine differ from, say, an intention of the form 'I intend to contribute to *J-ing*'? It is tempting to think that 'we *J*' as it appears in the content of my intention, must refer to a collective action, presupposing shared intention In that case, though, the account will not help us to a better understanding of what it means for an intention to be shared, and for an action to be collective It will not provide the meaning of "sharedness "

In his book *Complicity*, Kutz contrasts his own account of collective action with a reductionist view, according to which such acts can be seen as sums of individual choices "on the basis of expectations about the choices to be made by those with whom they wish to coordinate" (*ibid*, p 75) Kutz argues that this would be insufficient, and claims that "the core, minimalist, notion of collective action requires that individuals act on overlapping participatory intentions" (*ibid*) Implicit in the idea of a participatory intention is "a conception that one is doing one's part in a collective project" (*ibid*, p 76) Participatory intentions are a special class of intentions, "differentiated by their group-oriented content" (*ibid*, p 86) The content of this type of intention is "irreducibly collective" (*ibid*) So, the question is whether we can understand what kind of content these participatory intentions have without understanding the concept of a collective action in the first place [2]

II RESPONSES

Bratman's first response to the circularity challenge was to assume that we should appeal to act descriptions "characterized in cooperatively neutral ways "[3] My individual intention must refer to an act of the sort that may be cooperative but need not be The act description must satisfy the "behavioral conditions" for joint activities without presupposing "the attitudes essential to cooperative activity " "There is, for instance, a clear sense in which we can go to New York together or paint the house together without our activity being cooperative" (*ibid*) This initial response to the circularity challenge, Bratman claims, is similar to a common way of escaping circularity in the analysis of "intentional action " If we want to say that an action is performed intentionally when it is explained by the right kind of attitudes, we must "eschew appeal to attitudes that include in their content the very idea of intentional action Instead, we limit ourselves to contents that, at most, involve a concept of action that leaves it open whether the action is intentional "[4]

---

[2] Margaret Gilbert's theory of collective action is a third example of an analysis that is open to a circularity challenge of the kind discussed here "What is key is the incorporation of an holistic notion of joint commitment The formulation [in *On Social Facts*, p 198] correctly implies that the parties to any joint commitment must possess this notion"—"The Structure of the Social Atom Joint Commitment as the Foundation of Human Social Behavior," in Frederic Schmitt, ed, *Socializing Metaphysics* (Lanham, MD Rowman and Littlefield, 2003), pp 39–65, on p 53 So, the individual states in which the notion of joint commitment must occur are also necessary conditions for a joint commitment to obtain Since her approach differs in important aspects from Bratman's and Kutz's (she regards the enterprise as essentially normative, and it not evident that her account of how joint commitments are created should be read as a conceptual analysis), it would reach too far to give a fair treatment of her view here

[3] Bratman, "Shared Cooperative Activity," p 97

Margaret Gilbert objects to Bratman that it is unclear how he construes 'that we *J*' in 'I intend that we *J*' She suspects, though, that "he has an individualistic construal in mind Consider his central case of painting the house together I shall take it that he construes my intending that we paint the house together as something like this my intending that the house is painted by virtue of my actions along with complimentary actions of yours "[5] In that reading, though, the analysis will fail to capture intuitions about a stronger sense of "sharedness" or "collectivity" in action, which prompted the need for revising standard theory of action to begin with

Gilbert's objection might seem unfair in the light of Bratman's further restrictions on shared intention, which require that interdependence and interlocking subplans also figure in the participant's conception of 'our *J*-ing' In terms of the full necessary and sufficient conditions of Bratman's analysis, we have a shared intention to smash the window if and only if (1) (a) I intend that we smash the window and (b) you intend that we smash the window, and (2) I intend that we smash the window in accordance with and because of (1a), (1b), and meshing subplans of (1a) and (1b), and you intend similarly, and (3), (1) and (2) are common knowledge between us

However, I think that Bratman's initial strategy against circularity nevertheless makes Gilbert's doubts reasonable According to that strategy, it must be left open whether the 'we smash the window', as it occurs in the content of your intention and my intention, is a case of shared intention Suppose I want the window smashed When I note your presence on the street, I think that if you act in a certain way, the window can be smashed as a result of both our acts, and I form an intention accordingly What I intend in that case is merely to get the window smashed, while predicting that your actions will be components in the process leading to that result This prediction may rest upon my knowledge that your intentions are similar to mine, and that our subplans are likely to mesh in a way that enables me to reach my goal There is mutuality and interdependence, in line with Bratman's requirements Still, I would say, nothing in this picture captures "sharedness" or "collectivity" in any sense distinct from what we can construe in terms of standard individualistic theory of action What we have is a complex set of individual intentions, beliefs about other people's intentions, and means-ends reasoning

---

[4] Bratman, "I Intend That We *J*," p 147

[5] Gilbert, *Sociality and Responsibility* (Lanham, MD Rowman and Littlefield, 2000) p 156

In other words, if the characterization of a shared intention allows the individual intentions that make it up to be directed towards a mere set of actions that happen to be grouped together in a purely behavioral sense, the distinction between a shared intention and a set of individual noncooperative intentions will be undermined The additional conditions do not suffice to ground collectivity (It is even more apparent that this strategy would not work in Kutz's case, since he explicitly states that additional conditions such as mutual knowledge, interdependent subplans, and so on, are superfluous)

Bratman's analogy between avoidance of circularity about "intentional action" and about "collective action" is somewhat misleading Collective and individual intentional actions both belong to the species of "intentional action " The present way of drawing the distinction between them presupposes that the intentions that explain an act (in a way that makes the act intentional) contain a reference to the acting unit itself—the individual or the collective Without presupposing any references to intention, it seems unproblematic to conceive of an individual acting unit, and to imagine that sort of conception placed in the content of an intention to act individually The trouble here is that if shared intentions are supposed to be what ties members together in a collective unit, it seems unavoidable that we refer to those very intentions when we characterize the conception of the acting unit that is supposed to figure in their distinctive content The natural solution working for intentional action—to avoid reference to attitudes involving the concept to be defined—has a higher price when applied to "collective action "

So, Bratman's first solution seems to make the theory more reductionist than it was supposed to be With explicit reference to an earlier work by Kutz, he also appears to acknowledge this in the later paper [6] This brings us to the *second response* In Bratman's terms, that alternative strategy is to "seek basic cases in which each participant intends a joint activity understood in a way that is neutral with respect to shared intentionality. We then try to understand those cases in which the agent's conception of the shared activity explicitly supposes it is a shared activity, in a way that builds on our understanding of these basic cases" (*ibid.*, p. 148).

---

[6] Bratman, "I Intend That We *J*," pp 147–48

Kutz notes that there are cases where it seems proper to say that an intention is directed towards a jointly intentional activity. "[T]he intentions to perform many kinds of jointly intentional acts cannot be made sense of except in collective terms: noncooperative chess is not chess but something else—mutual chess solitaire" (*op. cit.*, p. 86). "[T]here is no way to characterize the intention of an individual planning to dance the tango except as an intention to engage in jointly intentional tango dancing" (*ibid.*). The proper response to the circularity challenge, according to Kutz, is therefore not "an analysis that tries to show how each instance of collective action is built out of non-collective materials, but rather a genealogical account that shows generally how the capacity to engage in collective action emerges out of capacities explicable without reference to collective concepts" (*ibid.*). This article discusses circularity in the definition of "collective action." This second response by Kutz and Bratman indicates that what they have in mind could be another issue—about circularity in the explanation of collective actions. What we need, in Kutz's words, is "a genealogical account that shows generally how the capacity to engage in collective action emerges out of capacities explicable without reference to collective concepts" (*ibid.*).[7] Roughly, it says that the content of an executive intention (that is, an intention regarding an outcome as a whole, whose function is to generate and command other subsidiary intentions) can come to refer to a jointly intentional act through a kind of bootstrapping interaction with simpler noncollective intentions, which then gradually become subsidiary to it:

> In learning to play chess, I first conceive of the elements of the game in nonjoint terms: the knight moves so and so, it is best to open with pawns, and if you threaten my queen, I should check your king. As I learn to play, these constituent elements of chess come to be represented and internalized as "my playing chess," or, alternatively, "my doing my part of our playing chess together." By such a bootstrapping process, the

---

[7] A similar genealogical response to the circularity challenge, along with the claim that the ensuing conception of jointness should be regarded as primitive, is suggested by Frederic Schmitt (on behalf of Gilbert) in the introduction to *Socializing Metaphysics*, p. 11.

> collective joint activity thus becomes the object of an agent's executive intention, having been built out of noncollective elements.[8]

This response meets one sort of circularity worry, by describing how an intention directed toward a jointly intentional activity can emerge from other sorts of intentions. One could imagine a similar response to Hume's circularity insinuation against the idea that a desire to act justly is what explains a just act.[9] The reply to Hume would then perhaps describe how the desire to perform a just action emerges out of purely self-interested desires. That would show how one might explain the desire in question, and thereby the performance of a just act, in a noncircular way. *Explanans* does not have to contain *explanandum* just because the explanation refers to an attitude in the content of which a conception of *explanandum* figures. Assume, though, that Hume's argument should instead be understood as a claim about circularity in definition, directed towards an analysis of the following sort. A 'just act' should be understood as 'an act performed out of the right kind of motive', and 'the right kind of motive' means 'a desire to act justly'. That would be an analytically circular definition, and the genealogical response would not remove its circularity. If the definition is supposed to provide the meaning of 'just act', it fails. In order to understand what it is that a person comes to desire when a desire to act justly emerges, we would have to know what a just act is to begin with.

The challenge to be met here is that in order to make clear what it means for an intention to have an "irreducibly collective content," we must avoid references to collective action. Otherwise, we cannot avoid circularity in the definition of the latter concept. Kutz's description could be regarded as a response to the accusation that *explanans* contains *explanandum*. It might also be read as a defense of the individualistic presupposition that the capacity to act collectively cannot be just as basic as the capacity to act individually. The account above describes how intentions directed towards jointly intentional activities can emerge, on that presupposition. But the initial circularity challenge is neutral with respect to this individualistic assumption, as well as to the question about circularity in explanation. The challenge is not that intentions with a noncollective content must come before intentions with a collective content, but that we need a characterization of the content of the components referred to in the definition of a collective action—a characterization that does not rely on that very concept This requirement is reasonable independently of which genesis we assume that these components have As far as I can see, Kutz's strategy fails to meet that challenge Even if his story above is correct, we still want to know what makes an action collective What kind of entity is it that becomes the object of the agent's intention through the bootstrapping process? His answer, like Bratman's, seems to be that the action turns collective when the executive intentions of the individuals contributing to it become directed towards a jointly intentional achievement That answer brings us back to the original circularity challenge, independently of how we think that these executive intentions come into existence

---

[8] Kutz, *Complicity*, pp. 87–88.

[9] Hume, *A Treatise of Human Nature*, book III, part II, section I.

A *third response* to the challenge would be to argue that the analysis is sufficiently informative in spite of its being circular Kutz hints at that strategy when he claims that the circularity problem is "more methodological than substantive " He also notes approvingly that Bratman has "relaxed his strictures" in this matter (*ibid.*) However, neither Bratman nor Kutz makes clear in which sense and to which extent the strictures concerning circularity should be relaxed in this specific context

Circular analyses may be informative and explanatory Philosophers have offered nonreductive explications of (for example) normative and modal concepts, while presupposing some primitive notion of modality or normativity That sort of analysis might be the best we can do in some cases Network definitions that place a concept within a larger scheme of interrelated concepts can be informative even if each concept is defined entirely in terms of others within the scheme And so on

One thing that may make some inclined to accept circularity in the present case is that the analysis merely suffers from what I L Humberstone calls analytical circularity It is not *inferentially circular*[10] Inferential circularity is generally vicious, while analytical circularity "might not always be a bad thing" according to Humberstone (*ibid*, p 254) Suppose we claim that a necessary condition for 'I smash the window intentionally' to be true is that I know that I smash the window intentionally (and that knowledge presupposes truth) This would be a case of inferential circularity If you know that this specific component of the analysis applies to a case, you already know that the concept analyzed applies You cannot know that I know that I smash the window intentionally unless you know that I smash it intentionally So, the analysis would neither enhance our understanding of the meaning of the term "intentional action," nor improve our ability to apply the term to the right kind of objects But suppose we want to analyze the notion of acting collectively, in the strong sense of "acting with a shared intention " We assume that a necessary condition for 'we smashed the window collectively' to be true is that our smashing the window was explained by my intending that we smash the window and your intending that we smash the window (The two latter occurrences of 'we smash the window' should be taken as presupposing shared intention) In that case, we still have analytical circularity—the concept analyzed appears (implicitly) in a component of the analysis However, this would not suffice to make the analysis inferentially circular, since we could know that the conception of a collective action, involving the notion of shared intention, appeared in the contents of your intention, without knowing that 'we smashed the window collectively' applies to the case As Humberstone puts it, the concept analyzed has a suitably protective embedding in this specific condition

---

[10] Humberstone, "Two Types of Circularity," *Philosophy and Phenomenological Research*, LVII, 2 (1997) 249-80

Analytically circular analyses often occur in contexts where we have a common pre-theoretical understanding of the concept analyzed, but a controversy or uncertainty concerning the applicability of the term to some given type of item In such cases, we may regard the list of application conditions "not so much as a way of getting the concept across to someone not familiar with it, but as a recipe for telling us when it applies" (*ibid*, p 251)

This possible excuse for analytical circularity does not apply here "Collective action" is not alleged to be like "art"—a concept of which we have a common pre-theoretical understanding while we sometimes need to be informed about boarders for application On the contrary, the starting point for analyses of "collective action" is an assumption about what kind of entities this concept applies to Proponents of the concept (like Bratman, Gilbert, and Kutz) typically focus salient cases of cooperative, joint action, as opposed to mere parallel behavior The analytical task they set themselves is to make clear how the description of these phenomena should be fitted into some overall action-theoretical framework

As Humberstone and Roseanna Keefe both stress, the purpose of the analysis determines whether we should accept its circularity[11] Circularity should at least not be accepted when "it obstructs the transfer of understanding the application conditions of a concept may be designed to effect "[12]

---

[11] Keefe, "When Does Circularity Matter?" *Proceedings from the Aristotelian Society*, cii (2002) 253–70

One sort of understanding that a definition of "collective action" should provide is practical Although the theories discussed here are descriptive, they will have relevance in moral and legal contexts (Kutz explicitly builds his theory about legal and moral accountability on his analysis of "collective action ") To assess normative claims about group actions and participation, we need to know what it is that the individuals are supposed to intend or be committed to when a collective action is said to emerge from their attitudes Since the circularity discussed here makes it unclear what kind of objects these attitudes have, it obstructs the transfer of one important kind of understanding that these analyses are designed to effect

I suppose that the main theoretical purpose of the analyses at hand is to establish a substantial distinction between collective actions and noncollective sets of individual actions Circularity clearly threatens that aim So, in this case, analytical circularity is a methodological problem, and a substantive one

III COLLECTIVITY WITHOUT INTENTION

Bratman suggested initially that circularity could be avoided by appealing to act descriptions that satisfy the behavioral conditions for joint activities but do not presuppose shared intention On that assumption, there is no reference to shared intention in the notion of joint activity that is supposed to figure in the content of the individual 'I intend that we *J*'-intention My complaint was that if we skip that implicit reference, Bratman's theory does not uphold the distinction between a collective action, and a set of individual interdependent actions The additional conditions about mutual interdependence, and the like, are insufficient for that

However, I believe that Bratman's suggestion could be elaborated An act description might be neutral with respect to shared intention without being neutral about collectivity The strategy is to employ a broad concept of collective activity that does not presuppose shared intention but is nevertheless substantively distinct from the notion of a mere set of intertwined acts That might require the description to go beyond mere behavior, though not very far

Susan Hurley employs a useful notion of a "unit of activity "[13] She does not offer any precise explicit definition, but develops the conception by utilizing it in a series of arguments directed towards an individualistic presupposition about instrumental rationality An important point is that the root of problems of coordination and cooperation (exemplified, for example, by prisoner's dilemmas) is not "the nature of the individuals' goals, or the instrumental character of rationality Rather it is individualism about rationality, which holds the unit of activity exogenously fixed at the individual" (*ibid*, p 201) Like Philip Pettit and others, Hurley thinks that considerations from game theory and social choice theory suggest that we must allow collectives to be regarded as agents or units of activity She denies, though, that this must involve regarding the unit of activity as the source of the evaluation of outcomes "Thus, collective activity does not require collective goals or preferences That is, collective action does not require that individual preferences be shared or even aggregated into some collective preference" (*ibid*, p 202) No joint goals, preferences, or intentions are required

---

[12] Humberstone, "Two Types of Circularity," p 251

[13] Hurley, "Rational Agency, Cooperation, and Mind-Reading," in Natalie Gold, ed , *Teamwork, Multi-Disciplinary Perspectives* (New York Palgrave, 2004), pp 200–15

If a definition of "unit of activity" could be extracted from Hurley's account, it would be purely operational Roughly, a unit of activity is an entity for which options can be evaluated instrumentally, at least from a spectator's point of view "When I speak of a unit of activity, I mean that activity within certain boundaries (which may be defined in spatial or other terms) has consequences, calculated against a background of what happens and is done outside those boundaries, the consequences of this activity can be evaluated instrumentally, in light of the goals of the unit or parts of it" (*ibid*, p 201)

Now, if a collective activity does not require shared preferences, goals, or intentions, where then is the glue that keep the participants together in a unit?[14] To frame the question differently, are there any essential intrinsic features that distinguishes collective units of activity from mere sets of agents? How do we distinguish a member from a nonmember?

Hurley offers several suggestions about how collective units of activity can be formed, and about how their boundaries become determined One approach focuses "the capacity to identify with others as part of a collective unit of activity" (*ibid*, p 208) The ability to attribute mental states to others plays an important role for the development of that capacity, Hurley argues However, "each participant's identification with the others as part of a collective unit of activity" cannot be taken as conceptually essential to a unit of activity

---

[14] This was Bratman's rhetorical question at Hurley's presentation of the paper I mainly refer to here (Collective Intentionality IV, Siena 2004)

To begin with, this would actualize a circularity challenge Furthermore, "unit of activity," as Hurley employs the term, applies to entities apparently incapable of identification, at least of the sort requiring the possession of intentionalistic concepts Slime mould can form units of activity "An ant colony can be regarded as a unit of activity governed by various local chemical signals that determine the behavior of individual ants" (*ibid*, p 206) It seems compatible with Hurley's use of the term to regard any entity as a unit of activity as long as it produces consequences such that it makes sense for a spectator to evaluate these effects in terms of goals assigned to the entity

Hurley stresses that different units of activity have different causal powers Attributing such powers to an entity appears to be a condition for regarding the entity as having options to begin with My suggestion is that that the problem of what binds a unit of activity together could be solved in terms of causal powers or dispositional properties This solution is based on the assumption that assignments of dispositional properties carry with them defining limits of the object to which these properties are assigned

To regard an entity as having causal powers or dispositions is to view that entity as a *causal agent*[15] This notion of an agent is even weaker than Hurley's "unit of activity " Assignments of purely causal agency do not presuppose that any goals can be assigned to the agent in question All causal agents are not units of activity in Hurley's sense This use of the terms "agent" and "agency" might seem linguistically inappropriate to some Perhaps they would say, with Donald Davidson, that agency by definition is intentional under some description "a man is the agent of an act if what he does can be described under an aspect that makes it intentional "[16] "I am the agent if I spill the coffee meaning to spill the tea, but not if you jiggle my hand" (*ibid*, pp 45–46) My spilling the contents of the cup is intentional, and that aspect is what makes me an agent of (unintentionally) spilling coffee

However, we sometimes categorize as acts reflexes, twitches, manifestations of common clumsiness, and other types of behavior, which are not intended under any description They fall under the notion of agency as different types of unintentional activities Under the adjective 'reflex', the *Oxford English Dictionary* lists 'reflex action (independent of the will, caused as automatic response to nerve-stimulation)' and the noun 'reflex' is said to be synonymous with 'reflex action' When I unknowingly talk loud to myself, or when I sneeze, you may think of me as *doing* something How, then, do these sorts of acts differ from things that merely happen to me, like when my hair grows or the earth beneath me moves me away from the sun? We cannot appeal to the empirical fact that I cannot choose to influence the latter sorts of events [17] That inability is part of the story about reflexes as well These are acts I cannot help doing

---

[15] Tim Crane employs this term when he exposes his view that dispositions are independent causes See Crane, "The Efficacy of Content," in Jan Bransen and Stefan E Cuypers, eds, *Human Action, Deliberation and Causation* (Boston Kluwer, 1998), pp 199–225, on p 220

[16] Davidson, "Agency" (1971) in his *Essays on Actions and Events* (New York Oxford, 1980), pp 43–61, on p 46

Verbs such as 'grow' or 'move' can be used to refer to processes or tendencies, or to activities A man's hair grows, and he may also grow a beard The ferry moves me away from the island, and I may actively move away from things Growing a beard and moving away from things can be actions that are intentional under some description However, that active sense of 'grow' or 'move' does not *entail* the presence of any intention A plant grows, but it is also capable of growing a new bud, as well as of moving its leaves towards the sun Without resorting to anthropomorphism, we might even explicitly say about a chemical, that it acts upon a certain substance by making it corrode or dissolve, and we may point to explanatory conditions as being *responsible* for their effects These everyday expressions indicate that ordinary speakers are capable of distinguishing a weak notion of agency, which is not intentional under any description This way of speaking simply places the object in a certain causal role, and refers to an effect for which internal features of the object is a condition

The notion of agency we employ when we talk about reflexes and other types of human behavior without intentional aspects does not only bear a metaphorical resemblance to the causal notion characterized above If you point out that your jiggling made *me* spill coffee, my causal role for that effect is stressed—although no intentions are in play My spilling coffee is in this case something that I do, albeit unintentionally, so it is an act, but not in a stronger sense than the one outlined The weak causal assumptions made about me in this case differ widely from the conditional statements that would have to be true in order for me to intentionally spill coffee

---

[17] This is Irving Thalberg's suggestion about how to distinguish "action verbs" from "bodily process and reaction verbs", see *Enigmas of Agency* (London Allen and Unwin, 1972), p 62 For more detailed criticism of this suggestion, see I Persson, *Reasons and Reason-governed Actions* (Lund Studentlitteratur, 1981), p 15

Although the causal notion of acting is weaker than the concept of "acting that is intentional under some description," there is a substantial difference between descriptions in terms of causal agency and purely behavioral judgments. The latter are even less committing. The coffee incident could be described in purely behavioral terms—your hand comes in contact with my wrist, my hand trembles, this causes coffee to be spilt, and so on—and that sort of description would be neutral with respect to causal agency. Similarly, we could describe the series of events that occur when the liquid and the metal meet, while leaving the question of causal agency open. A description of our going to New York, or our painting the house, which merely satisfies the behavioral conditions for these activities, would in itself contain no commitments concerning causal agency. Connected with the use of a broad merely causal notion of agency is a claim about dispositional properties. To say that the chemical acts upon a substance is to assume that there is something about the constitution of the chemical, which explains what happens to the substance under the actual conditions. 'The plant moves its leaves' (as compared to 'the plant's leaves move') contains a similar allegation about the constitution of the plant. If you stress that I was the one who spilled coffee, albeit unintentionally, you are implicitly claiming that there is something about my features, which explain why I spilled coffee under those conditions. This sort of claim is closely associated with the assignment of a dispositional property to the object in question—in the last case, perhaps clumsiness or feebleness.

When we explain what happens with the metal in terms of what the acid does to it, and thereby place the liquid in the role of the causal agent, we presuppose a specific level of explanation. If we were chemists, we could have chosen to describe the process in terms of what the molecules do. We would then assign dispositional properties on the molecular level. Presumably, molecular dispositions form the base of the acid's disposition of being corrosive. That does not imply conceptual reduction of descriptions of the latter to descriptions of the former. Even a chemist in the habit of thinking about chemical processes in molecular terms switches easily to macro-level descriptions in contexts where it is linguistically inappropriate to go into molecular details when talking or thinking about what a certain chemical can do.

Our choice of a certain object as the causal agent implicitly involves defining the unit to which causal powers are assigned (in the sense of drawing a boundary, not in the semantic sense of 'define'). The object's components belong to it in virtue of their roles for the object's intrinsic dispositional properties or causal powers. To regard a set of objects as one causal agent is to imply that *something* glues its components together, without being committed to the attribution of some specific glue.

To illustrate my claim that disposition-assignments implicitly define the unit to which causal powers are attributed, let me use an example from the philosophical debate on the nature of dispositions. George Molnar presents the existence of "intrinsic maskers" of dispositions as a case against the idea that disposition-assignments can be reduced to conditional causal statements of a certain kind. According to Molnar, an object *x* may have a disposition *D*, and at the same time there may be another disposition intrinsic to *x*, *D**, such that its being triggered prevents the manifestation of *D*. Intrinsically masked dispositions, like the "power of ingested poison masked by the power of ingested antidote," are "everywhere once you start looking for them." (So, Molnar claims, we cannot reduce assignments of dispositions to talk about conditional predictions of their manifestations.)[18]

It is essential to Molnar's argument that the masking and the masked disposition are intrinsic to one and the same object. The assumption that they are intrinsic presupposes some boundary around whatever they are intrinsic to. Suppose substance *y* makes us tired. In the drug *x* this is outweighed by the effect of an additional substance *z*. We could say with Molnar that *x* has two dispositional properties. It is fatiguing, and it has the property of blocking manifestations of the first type of disposition. The drug then has two causal powers that happen to cancel each other out. A causal conditional analysis of dispositions would not admit that *x* possesses both these dispositions. On the other hand, we might deny that *x* has any dispositional properties of the kind mentioned. We could instead assign the property of being fatiguing to substance *y*, which is an element of *x*. That would be compatible with a (sufficiently subtle) conditional causal understanding of the disposition in question.[19]

My point is not that any of these two ways of regarding the matter would be more plausible than the other, but that they exploit different underlying assumptions about the individuation of the bearer of the dispositional property in question, and that any assignment of causal powers or dispositional properties comes with such implicit assumptions. Our responses to the "intrinsic maskers" problem will not only depend on our intuitions about the conditional analysis of dispositions, but also on our ways of determining the boundaries of their bearers.[20]

---

[18] Molnar, "Are Dispositions Reducible?" *The Philosophical Quarterly*, XLIX (1999): 1–17, p. 5.

[19] The "intrinsic maskers" example is meant to undermine even a subtle version of the conditional analysis, of the type proposed by David Lewis. See "Finkish Dispositions," *The Philosophical Quarterly*, XLVII (1997): 143–58. Lewis's analysis allows for the possibility of *extrinsic* maskers. If we assign the property of being fatiguing to substance *y*, and the property of blocking the manifestations of that disposition to substance *z*, the latter disposition would be an extrinsic masker of the first.

How much structure do we require from a set of entities before we start placing them within the boundaries of one causal agent? Vice versa, under what circumstances will we switch from viewing an object as one causal agent to regarding it as a set of distinct but perhaps interdependent causal agents? These are questions of pragmatics. Their answers depend on which level of assignment of causal agency that is most convenient for explanatory or predictive purposes in the context. Someone familiar with ecology may find it natural to think about what one species does to another, or to the ecosystem, while others would tend to regard a species of plants or animals as nothing but a loosely connected set of distinct units of agency. On the other hand, when it comes to saliently demarcated macroscopic objects, like a spot of acid on a metal plate, few of us find it useful to regard that as a mere conglomerate of causal agents. Only certain spectators with specific background knowledge and training do that, and only in special circumstances.

There are also cases such that most of us switch easily between the two perspectives. Ant colonies or swarms of bees would be paradigmatic examples. Evidently we are capable of thinking about a swarm of bees as a set of interdependent but distinct causal agents. We may also watch the swarm as one causal agent and think about what it might do, wonder what makes it fly this way rather than that, think that some of its acts seem unexpected, and so on. That way of thinking does not require much knowledge about what keeps the swarm together. You may think that it has something to do with chemical signals while the small child beside you has a completely different theory. What makes both of you regard the swarm as the unit of causal agency need not be any specific knowledge about its internal structure, but just that its behavior gives you reasons to think that there is some such structure.

---

[20] Dispositions are the subject of an old but still ongoing metaphysical debate and few if any claims about their nature are uncontroversial. Two claims relating to the concept of dispositions are essential to my argument here. First, that regarding an entity as a causal agent—like when we think of the acid as acting upon the metal—is closely associated with assigning dispositional properties to the entity. Second, that assignments of dispositional properties carry with them defining limits of the entities to which they are assigned. These are claims about commitments inherent in ordinary people's thinking and talking. My purpose here is to explicate a certain conception of causal agency that may figure in the content of ordinary people's attitudes. Although I think that these two assumptions are compatible with quite differing accounts of the nature of dispositions, they may not square with any such account. To explore the exact metaphysical implications of these proposals would reach beyond the ambitions of this paper, though.

In a similar manner, we are capable of changing perspective when we regard the behavior of people. Without invoking any concept of intention, we may distinguish between the behavioral disposition of a group of agents, and the set of behavioral dispositions of its members. A group of people may be disposed to smash a shop window under certain conditions. The base of this disposition could be a set of behavioral dispositions of the group's members. This does not imply, though, that the behavioral disposition of the group is conceptually reducible to that set of individual dispositions. In other words, we can regard a group of individuals as the unit of causal agency, in a sense clearly distinct from that in which we view it as a set of causal agents. My contention is that to think of a set of individual actions as a *collective activity* is to regard the set as the causal agent. In regarding the group as the causal agent, we imply that there is some glue—there is something about the intrinsic features of the group and about the participants' role in the base of the group's causal powers, which distinguishes members from nonmembers—although we refrain from specifying this glue. This way of implying that there are internal features of an object responsible for the object's behavior, without specifying those internal features, is characteristic of any kind of assignment of dispositional properties. Typically, assignments of dispositions are informative and explanatorily valuable in cases where it is nontrivial that the explanation of an object's behavior is to be found among its internal features, although it would be impractical to go into the details of those features (because they are unknown, or it would take too long time, or be too costly for some other reason).

The weak notion of collective activity, in which the group is conceived of as the causal agent, is not an ad hoc construction, designed to save a certain type of account of collective action from circularity. There are, I think, independent reasons to regard this way of thinking about groups as part of common sense. Apparently, we have the capacity to lump things like bees together in a way that enables us to regard sets of them as distinct units of causal agency. I see no reason to think that this conceptual ability must make an exception when it comes to sets of people.

One sort of indication of our possession of a concept of collectivity without intention is that there is some empirical evidence for ascribing a conception of collectivity to two-year-old children.[21] There is also well-known evidence strongly suggesting that the capacity to assign intentional states (beliefs, desires, intentions) to others as well as to ourselves is developed later in life—usually around four years.[22] If both findings are correct, they prove that two-year-old children have a conception of collectivity that does not presuppose intention. However, experimenters' descriptions are unavoidably theory laden and empirical reports on issues like these are usually open to interpretation in accordance with more than one conceptual scheme. When it comes to the claim that the weak notion of collectivity is an element in ordinary people's way of thinking and talking, I will in the end have to rely on appeal to linguistic intuitions.

Suppose you are engaged in a discussion while walking with a small group of people. After a while, you all realize that you have walked past the place you were supposed to stop at. Each of you just automatically followed the others without thinking about where you were heading. None of you intended to walk that far, let alone that the group should do so. Still, it seems plausible to regard your walking behavior as collective, in a sense distinct from a set of individual acts of walking that far. I would regard this as a typical example of a collective activity where no intentions concerning the collective activity (in this case, the activity of walking that far) bind the participants together. The analysis suggested here provides a way of understanding how we can think of such activities as being collective in a distinct sense.

IV. CONCLUSION

When someone is watching us smashing the shop window, he can think of what we are doing as a *collective activity*—a result of the group's disposition being triggered. Then he simply thinks of the window's being smashed as explained by features of the group as a whole, and the group as the unit of causal agency. In other words, he can regard our behavior as collective, in a sense that does not presuppose shared intention. If each of us intends that the group smashes the window, and conceives of the intended activity as collective, that is, of the set of agents as the unit of causal agency in what is intended, then we collectively intend to smash the window. (Here I assume that conditions of mutual knowledge and other requirements of the preferred analysis are fulfilled I leave it open which additional conditions, if any, that are needed) To use a handy hypostatization, we might say that this set of I-intend-that-we intentions is a collective intention Insofar as this collective intention is what explains our smashing the window, it is a *collective action* We do not need to place the notion of a jointly intentional achievement in the content of the intentions of the parties to a collective action The weaker notion of collective activity is sufficient for that role

---

[21] This empirical research was presented by Hannes Rakoczy, Max-Planck-Institute for Evolutionary Anthropology, at Collective Intentionality IV, Siena 2004 ("Pretend Play and the Development of Collective Intentionality," work in progress).

[22] See, for example, A. Gopnik, "How We Know Our Minds: The Illusion of First-Person Knowledge of Intentionality," *Behavioral and Brain Sciences*, xvi (1993): 1–15, 90–101. (I owe this reference to Lars Hall and Petter Johansson.)

Take a paradigm case of collective action A good soccer player thinks partly in terms of how his or her team should act, not merely in terms of individual achievement Suppose each player at a certain moment in a game, perhaps at a signal from the coach, starts thinking that the best thing for the team to do would be to settle with the present score and go for mere defense They should not attempt to score more goals at the price of risking the points they have at present Each player then comes to intend that the team becomes more defensive That requires each of them to think about the team in a way that concerns the team's behavioral dispositions This is the most convenient way of thinking about this change of strategy They would think of the team as the unit of agency, and evaluate the options that lie within that unit's causal powers I guess most of us can recognize this simple switch between an individual and a collective perspective on the behavior of a group to which we belong It seems farfetched to require that the players conceive of their intended change of strategy as being like tango dancing, that is, an activity for which it is essential that each participant conceives of what they are doing as a jointly intentional activity Occasionally thinking in that way about each other's engagement might be important for getting a greater enjoyment out of playing soccer, compared to the pleasure of playing computer games or solitaire, for instance Simple group actions within the game do not require that sort of self-reflective awareness, though

My claim is that a weak notion of collective activity must figure in the content of the intentions of the parties to a collective action That claim is nevertheless stronger than Bratman's initial proposal that it is sufficient that the act description, as it figures in the content of the agent's intention, merely satisfies the behavioral conditions for a joint activity Such a description would be neutral with respect to causal agency (A behavioral description like 'their arms move' makes no allegations about causal agency) My additional requirement is that the notion of the team as the causal agent must enter that content It might even be misleading to regard this requirement as appealing to intentions "characterized in cooperatively neutral ways," in line with Bratman's initial proposal[23] If we allow for a weak sense of cooperation, corresponding to the notion of causal agency, ants, slime mould and molecules can be said to cooperate in relation to the effects of the unit of causal agency to which they belong The implicit assumption about there being something that binds the elements together could, as far as I can see, be expressed in terms of "causal cooperation "

Finally, I should make it clear that I do not deny that participants in some groups actually think of their intended common activity as being jointly intentional, that is, that they intend their collective activity to be a collective action Each participant may intend that each participant intends the activity to be collective It is essential to some individual action-types, like voting, that they are carried out intentionally When I intend to vote, I conceive of this act as intentional In a similar manner, it may be essential to some collective action-types, like tango dancing, that they are jointly intentional When we intend to dance a tango together, each of us conceives of our intended act as being jointly intentional There is no reason to exclude the possibility of self-reflexive capacities of that sort It is not implausible to think that such abilities may develop in the bootstrapping manner Bratman and Kutz suggest The important thing is that this higher-order capacity is inessential to the ability of performing a jointly intentional collective act—a collective action All collective actions do not have to be that sophisticated

BJÖRN PETERSSON

Lund University

---

[23] Bratman, "Shared Cooperative Activity," p 97

KNOWLEDGE AND EVIDENCE

In *Knowledge and Its Limits*, Timothy Williamson argues that knowledge is central in many ways.[1] For one thing, he argues that evidence is knowledge. And he also argues that there is a sense in which knowledge is conceptually prior to belief, that "the place of belief in the economy of the mental life depends on its potential connection with knowledge" (KL 95).

Consider the sorts of evidence deployed in standard accounts of the important case of perceptual beliefs. Experience provides one such crucial kind. Or if all evidence is propositional and experience is not, experience is at least closely related to one such kind. And so it is quite relevant that some of Williamson's key arguments are focused on experience.

It is natural to object to any equation of evidence and knowledge that we cannot introspect in some infallible way whether we have knowledge, but that we can introspect in that way, at least in central and crucial cases involving experience, what our evidence is. But Williamson provides an interesting argument that no nontrivial state of mind meets such a stringent condition of privileged access. It is my focus here.

Williamson says that a condition *E*, for instance, the nature of one's current experiences, is "luminous" iff in every situation, if in that situation *E* obtains, then in that situation one is in position to know that *E* obtains (KL 97). But he argues that no mental states, not even those to which we are customarily deemed to have privileged access, are luminous.

"Consider a morning on which one feels freezing cold at dawn, very slowly warms up, and feels hot by noon. One changes from feeling cold to not feeling cold, and from being in a position to know that one feels cold to not being in a position to know that one feels cold. If the condition that one feels cold is luminous, these changes are exactly simultaneous" (KL 97).[2] But he argues, they are not simultaneous (KL 97–98). The key to the argument is Williamson's claim that knowledge must be reliable, in a sense which implies that one would continue to know whether one now feels cold even in a situation in which one felt a trifle warmer or colder. I myself do not see why

---

[1] New York: Oxford, 2000, hereafter KL.

[2] This argument appeared earlier in Williamson, "Cognitive Homelessness," this JOURNAL, XCIII, 11 (November 1996): 554–73.

0022-362X/07/0403/157–60

perfectly precise introspective knowledge—introspective knowledge which captures such fine details that it would vary with any tiny detail in what is introspected—would be unreliable in an epistemically relevant sense But this claim is crucial to Williamson's argument that experience is not luminous, so I waive this objection Here is the way his argument works

Let *Ai* be one's mental state at millisecond *i* in the interval between dawn and noon, and *Ai*+1 be one's mental state in the following millisecond

(1) If in *Ai* one knows that one feels cold, then in *Ai*+1 one feels cold

This is assured by reliability

(2) If in *Ai* one feels cold, then in *Ai* one knows that one feels cold

This is assured by luminosity

Now suppose

(3) In *Ai* one feels cold

By modus ponens on (2) and (3), we get

(4) In *Ai* one knows that one feels cold

And modus ponens on (1) and (4) yields

(5) In *Ai*+1 one feels cold

But at dawn, in *A*0, one feels cold And we can run such an argument for that initial millisecond and then for every successive millisecond until noon, and conclude that at noon one feels cold, when indeed it is very hot. Because this argument depends merely on luminosity and reliability, Williamson claims that luminosity must go

While, as I said, I doubt that reliability in his sense has more antecedent plausibility than luminosity, my point here will be to explore the interaction of this argument with another of Williamson's key claims and arguments

Williamson admits that experience provides evidence, as he plausibly must (KL 197) Since he presumes that evidence is knowledge and hence is propositional, he thinks he cannot allow that experience *is* evidence, and so he claims that experience provides evidence by conferring the status of evidence on propositions (KL 197)

But what propositions? The answer is a sort familiar from the work of John McDowell[3]

---

[3] *Mind and World* (Cambridge Harvard 1996), pp 56–59

Here is a case "I am trying to identify a mountain by its shape I can see that it is pointed, that it is pointed may be part of my evidence for believing that it is not Ben Nevis However, the proposition that it is pointed does not begin to exhaust my present perceptual evidence No description of the mountain in words seems to capture the richness of my visual experience of its irregular shape" (KL 197) This may seem a problem for the claim that evidence is propositional and indeed knowledge, but there is a way out Williamson suggests "I might point and say, 'It is that shape'" (KL 198) Only in using that sentence in an appropriate context which includes such an experience does he express the proposition in question,[4] but such a use might constitute knowledge which is perceptual evidence properly responsive to experience

But remember that Williamson thinks that knowledge must be reliable, so it cannot be too exact It must be "unspecific enough" (KL 198)

There are three things the mountain, his experience of the mountain, and his knowledge The mountain is exact The knowledge is unspecific But does the play come between mountain and experience, or between experience and knowledge, or both?

Williamson claims that the unspecificity in his account makes it closer to the views of McDowell than of Christopher Peacocke, and gives references to discussions in which Peacocke allows and McDowell denies that there is any gap between thought and experience [5] So the alleged gap must be between experience and the world And what he says about the case seems to confirm this "Of course, I cannot see *exactly* what shape the mountain is, I can only see roughly what profile it presents to me, and cannot see round the back" (KL 198)

There is no gap of specificity between thought and experience, and there is such a gap between experience and world Apparently Williamson believes such conditions are required if he is to plausibly square his claims that evidence is knowledge and that knowledge requires reliability with the obvious fact that experience provides evidence

But consider again Williamson's argument against luminosity There he needed a gap between thought and experience, a gap which we now know he thinks can in fact be bridged by McDowellian perceptual demonstrations And there he needed there to be no relevant gap between world and experience, so that our feelings of cold can differ in the same indiscriminable way as differences in temperature, so that

---

[4] McDowell suggests that a past experience will do

[5] Peacocke, *A Study of Concepts* (Cambridge MIT 1992), pp 83–84, McDowell, pp 56–59, KL 198, note 9

NEW BOOKS ANTHOLOGIES

LACHS, JOHN, and D MICAH HESTER, eds *A William Ernest Hocking Reader* Nashville, TN Vanderbilt, 2004 xv, 400 p Cloth $95 00, paper $45 00

LACKEY, JENNIFER, and ERNEST SOSA, eds *The Epistemology of Testimony* New York Oxford, 2006 vi, 312 p Paper $35 00

LANGE, MARC, ed *Philosophy of Science* Malden, MA Blackwell, 2007 xiv, 649 p Paper $49 95

LEAK, ANDREW, and ADRIAN VAN DEN HOVEN, eds *Sartre Today A Centenary Celebration* New York Berghahn, 2005 xi, 330 p Paper $89 95

LILLEHAMMER, HALLVARD, and D H MELLOR, eds *Ramsey's Legacy* New York Oxford, 2005 x, 182 p Cloth $74 00

LINES, DAVID K , ed *Music Education for the New Millennium* Malden, MA Blackwell, 2005 vii, 150 p Paper $34 95

LINZEY, ANDREW, and PAUL BARRY CLARKE, eds *Animal Rights A Historical Anthology* New York Columbia, 2004 xxxi, 193 p Cloth $64 50, paper $24 50

LUDLOW, PETER, DANIEL STOLJAR, and YUJIN NAGASAWA, eds *There's Something about Mary Essays on Phenomenal Consciousness and Frank Jackson's Knowledge Argument* Cambridge MIT, 2005 xix, 463 p Paper $35 00

MACDONALD, GRAHAM, and DAVID PAPINEAU, eds *Teleosemantics* New York Oxford, 2006 vii, 232 p Paper $29 95

MACHAMER, PETER, and GEREON WOLTERS, eds *Science, Values, and Objectivity* Pittsburgh University Press, 2004 vii, 317 p Cloth $44 95, *Thinking about Causes From Greek Philosophy to Modern Physics* Pittsburgh University Press, 2007 vi, 318 p Cloth $44 95

MADIGAN, TIMOTHY J , ed *Promethean Love Paul Kurtz and the Humanistic Perspective on Love* Newcastle, UK. Cambridge Scholars, 2006 xii, 327 p Cloth $69 99

MANOUSSAKIS, JOHN PANTELEIMON, ed *After God Richard Kearney and the Religious Turn in Continental Philosophy* New York Fordham, 2006 xx, 439 p Cloth $85 00, paper $28 00

MARIÑA, JACQUELINE, ed *The Cambridge Companion to Friedrich Schleiermacher* New York Cambridge, 2005 xiv, 348 p Cloth $70 00, paper $29 99

MARMION, DECLAN, and MARY E HINES, eds *The Cambridge Companion to Karl Rahner* New York Cambridge, 2005 xvi, 318 p Cloth $70 00, paper $25 99

MARSO, LORI JO, and PATRICIA MOYNAGH, eds *Simone de Beauvoir's Political Thinking* Urbana Illinois UP, 2006 viii, 139 p Cloth $50 00, paper $18 00

MARTIN, MICHAEL, ed *The Cambridge Companion to Atheism* New York Cambridge, 2006 xix, 331 p Cloth $75 00, paper $27 99

MCGEE, GLENN, and ARTHUR CAPLAN, eds *The Human Cloning Debate 4th Edition* Berkeley, CA Berkeley Hills, 2004 vi, 330 p Paper $15 95

MCGILVRAY, JAMES, ed *The Cambridge Companion to Chomsky* New York Cambridge, 2005 ix, 335 p Cloth $75 00, paper $29 99

MCKENNA, ERIN, and ANDREW LIGHT, eds *Animal Pragmatism Rethinking Human-Nonhuman Relationships* Bloomington Indiana UP, 2004 xii, 254 p Cloth $50 00, paper $22 95

MENGUE, ANGUS J L , ed *Reading God's World The Scientific Vocation* Saint Louis, MO Concordia, 2004 336 p Paper $14 99

MERCER, CHRISTIA, and EILEEN O'NEIL, eds *Early Modern Philosophy Mind, Matter, and Metaphysics* New York Oxford, 2005 xxi, 298 p Cloth $55 00

MINKOW, SVETOZAR, and STÉPHANE DOUARD, eds *Enlightening Revolutions Essays in Honor of Ralph Lerner* Lanham, MD Lexington, 2006 xi, 399 p Cloth $90 00

MISAK, CHERYL, ed *New Pragmatism* New York Oxford, 2007 vii, 192 p Cloth $45 00

MOGGACH, DOUGLAS, ed *The New Hegelians Politics and Philosophy in the Hegelian School* New York Cambridge, 2006 xiv, 345 p Cloth $75 00

MOYAL-SHARROCK, DANIELE, and WILLIAM H BRENNER, eds *Readings of Wittgenstein's On Certainty* New York Global, 2005 xiii, 334 p Cloth $90 00

MUFWENE, SALIKOKO S, ELAINE J FRANCIS, and REBECCA S WHEELER, eds *Polymorphous Linguistics Jim McCawley's Legacy* Cambridge MIT, 2005 xxx, 550 p Cloth $85 00, paper $45 00

MYERS, SONDRA, and PATRICE BRODEUR, eds *The Pluralist Paradigm Democracy and Religion in the 21^{st} Century* Scranton, PA University Press, 2007 xiii, 165 p Paper $20 00

NARDIN, TERRY, and MELISSA S WILLIAMS, eds *Humanitarian Intervention Nomos XLVII* New York NYU, 2006 xii, 308 p Cloth $55 00

NEWMAN, LEX, ed *Locke's Essay Concerning Human Understanding* New York Cambridge, 2007 xiii, 486 p Cloth $80 00, paper $29 99

NICHOLS, SHAUN, ed *The Architecture of Imagination New Essays on Pretence, Possibility, and Fiction* New York Oxford, 2006 viii, 279 p Paper $35 00

NORRIS, ANDREW, ed *Politics, Metaphysics, and Death Essays on Giorgio Agamben's Homo Sacer* Durham, NC Duke, 2005 331 p Cloth $84 95, paper $23 95

O'CONNOR, BRIAN, and GEORG MOHR, eds *German Idealism An Anthology and Guide* Chicago University Press, 2006 vi, 481 p Cloth $95 00, paper $37 50

ODERBERG, DAVID S, ed *The Old New Logic Essays On the Philosophy of Fred Sommers* Cambridge MIT, 2005 xii, 242 p Cloth $75 00, paper $35 00

OLSSON, ERIK J, ed *Knowledge and Inquiry Essays on the Pragmatism of Isaac Levi* New York Cambridge, 2006 400 p Cloth $80 00

PEARSON, KEITH ANSELL, ed *A Companion to Nietzsche* Malden, MA Blackwell, 2006 xxii, 600 p Cloth $149 95

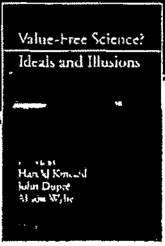
PERRY, JOHN, MICHAEL BRATMAN, and JOHN MARTIN FISCHER, eds *Introduction to Philosophy Classical and Contemporary Readings, 4th Edition* New York Oxford, 2007 xiv, 819 p Paper $72 00

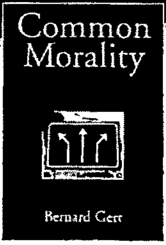
PETRAKIS, PETER A, and CECIL L EUBANKS, eds *Eric Voegelin's Dialogue with the Postmoderns Searching for Foundations* Columbia Missouri UP, 2004 ix, 187 p Cloth $37 50

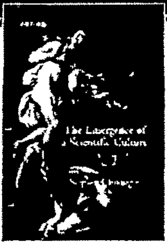
POLT, RICHARD, ed *Heidegger's Being and Time* Lanham, MD Rowman and Littlefield, 2005 viii, 244 p Cloth $75 00, paper $24 95

PRICE, HUW, and RICHARD CORRY, eds *Causation, Physics, and the Constitution of Reality Russell's Republic Revisited* New York Oxford, 2007 ix, 403 p Paper $35 00

PREYER, GERHARD, and GEORG PETER, eds *Contextualism in Philosophy Knowledge, Meaning, and Truth* New York Oxford, 2005 viii, 402 p Cloth $99 00, paper $35 00







New from OXFORD

New in paperback!
A Choice Outstanding Academic Title for 2005

The Philosophy of Psychiatry
A Companion
Edited by JENNIFER RADDEN

"The publication of this book is a major event in the ongoing development of the field . . . a significant accomplishment . . . it simultaneously represents, announces, and consolidates the arrival of an exciting new field."
—*Metapsychology Online Book Reviews (International Perspectives in Philosophy and Psychiatry)*

2004 (paper 2007) 480 pp. 1 line illus.
paper $35.00 cloth $74.00

New in paperback!

What's It All About?
Philosophy and the Meaning of Life
JULIAN BAGGINI

"Useful and provocative."—*Wall Street Journal*

"It's egalitarianism of style and content is admirable. There is nothing here to put off someone who has never read a book of philosophy, yet the book is doing philosophy, not just talking about it."—*Scotland on Sunday*

2007 224 pp. paper $12.95 cloth $23.00

The Emergence of a Scientific Culture
Science and the Shaping of Modernity 1210–1685
STEPHEN GAUKROGER

Why did science emerge in the West and how did scientific values come to be regarded as the yardstick for all other forms of knowledge? Stephen Gaukroger shows that science was bitterly contested during the early modern period. It did not distance itself from religion but rather entered into an alliance with it to provide a comprehensive picture of the world—transforming not just science but also religion.

2007 544 pp. $65.00

New in paperback!
A Choice Outstanding Academic Title for 2005

Common Morality
Deciding What to Do
BERNARD GERT

"I have admired and been excited by Bernard Gert's account of the moral system ever since I became acquainted with it. His account has made much good sense, and has seemed to correct much that has been problematic about past moral theories. Professor Gert's work is exciting because it presents a clear answer to one of the broadest questions in philosophy —What is the nature of morality?—and it does so in a way that has some of us beginning to think that, after all these centuries, someone has actually gotten it right."
—Timm Triplett, Associate Professor of Philosophy, University of New Hampshire

2004 (paper 2007) 192 pp. paper $15.95 cloth $29.95

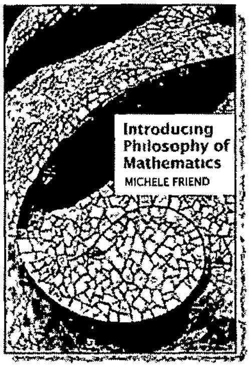
Value-Free Science
Ideal or Illusion?
Edited by HAROLD KINCAID, JOHN DUPRÉ, and ALISON WYLIE

It has long been thought that science is our best hope for realizing objective knowledge, but that, to deliver on this promise, it must be value free. Things are not so simple, however, as recent work in science studies makes clear. The contributors to this volume investigate where and how values are involved in science, and examine the implications of this involvement for ideals of objectivity.

2007 304 pp. $65.00

OXFORD
UNIVERSITY PRESS

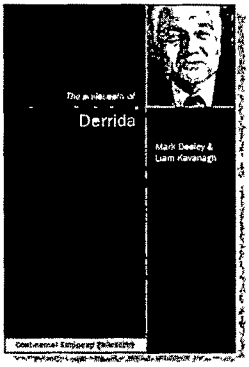
Prices are subject to change and apply only in the US. To order, please call 1-800-451-7556. In Canada, call 1-800-387-8020. Visit our web site at www.oup.com/us/



Introducing Philosophy of Mathematics
Michele Friend

Acumen Publishing
978 1-84465-061-3 $22.95 paper
978-1-84465-060-6 $85.00 cloth

A concise introduction to the standard theories of mathematics – platonism and realism, logicism, formalism, constructivism, and structuralism – as well as to some of the less standard theories, such as psychologism, fictionalism, and Meinongian philosophy of mathematics.

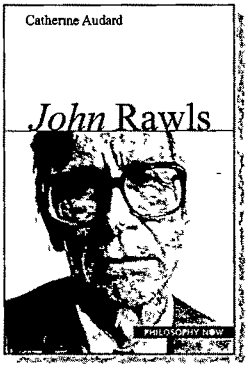


The Philosophy of Derrida
Mark Dooley & Liam Kavanagh

Continental European Philosophy
978-0-7735-3235-9 $22.95 paper
978-0 7735-3234-2 $80.00 cloth

"One of the most reliable and readable presentations of Derrida available. An engaging and provocative study that no reader of Derrida will want to miss."
JOHN D. CAPUTO, SYRACUSE UNIVERSITY

"Splendidly clear, lucid, and well-argued."
CHRISTINA HOWELLS, UNIVERSITY OF OXFORD



John Rawls
Catherine Audard

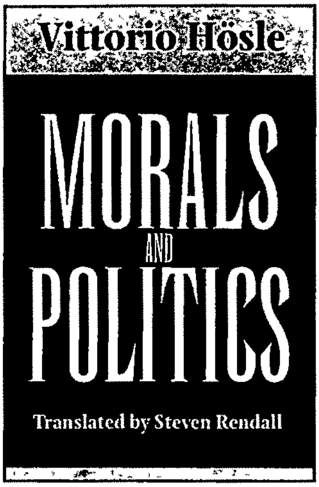
Philosophy Now
978-0-7735-3237-3 $22.95 paper
978-0-7735-3236-6 $75.00 cloth

"An impressive book by someone with an extensive and deep appreciation of Rawls' work. It offers a nuanced discussion of his central ideas."
MATTHEW CLAYTON, UNIVERSITY OF WARWICK

McGILL-QUEEN'S UNIVERSITY PRESS
www.mqup.ca

Now available in a new English translation . . .

"This book is impressive for its detail and sweep Various sections of the book will be useful for scholars in political science, political philosophy, and analytic sociology Recommended " —***Choice***



MORALS
AND
POLITICS
Vittorio Hösle
Translated by
Steven Rendall

"This is an intriguing and learned work on a huge scale In places this weighty tome reminded the reviewer of the intellectual challenge posed by reading Lonergan's *Insight* or Blondel's *L'Action* The author defends cogently his conviction that political philosophy must be based on ethics, while being aware that ethical arguments themselves have a political function Because of the vast scale of the topics covered and the great care in addressing each of them in detail, the consistency and coherence of the ethical approach to politics revealed here is exemplary in its cogency, at the same time as offering inspiration and guidance for politicians, policy makers, and indeed for all responsible citizens " —***Heythrop Journal***

ISBN 0-268-03065-0 • 978-0-268-03065-0 • $60 00 cloth • 1,016 pages

At bookstores, or order on the Web at. *www.undpress.nd edu*

UNIVERSITY OF NOTRE DAME PRESS
Chicago Distribution Center
11030 South Langley Ave , Chicago, IL 60628
Tel 800-621-2736 • Fax 800-621-8476

THE MONIST

AN INTERNATIONAL JOURNAL OF GENERAL PHILOSOPHICAL INQUIRY

JULY 2005 ◆ VOL 88, NO 3 ◆ SINGLE ISSUE, US $12

TIME TRAVEL

ADVISORY EDITOR ACHILLE C VARZI

ARTICLES	
ACHILLE C VARZI	*Foreword*
THEODORE SIDER	*Traveling in A- and B- Time*
ROBIN LE POIDEVIN	*The Cheshire Cat Problem and Other Spatial Obstacles to Backwards Time Travel*
JONATHAN SIMON	*Is Time Travel a Problem for the Three-Dimensionalist?*
MATTHEW H SLATER	*The Necessity of Time Travel (on Pain of Indeterminacy)*
PETER B M. VRANAS	*Do Cry Over Spilt Milk Possibly You Can Change the Past*
NICHOLAS J. J. SMITH	*Why Would Time Travellers Try to Kill Their Younger Selves?*
GORDON PARK STEVENSON	*Time Travel, Agency, and Nomic Constraint*
STEVEN SAVITT	*Time Travel and Becoming*
DAVID HORACEK	*Time Travel in Indeterministic Worlds*

Editor· Barry Smith, University at Buffalo, State University of New York

Single issues, $12 postpaid Annual subscription institutions, US $50, individuals, US $30, make checks payable to THE MONIST, 315 Fifth Street, Peru, IL U S A 61354

Website http://monist.buffalo edu E-mail philomon1@netscape net

---

THE MONIST

AN INTERNATIONAL JOURNAL OF GENERAL PHILOSOPHICAL INQUIRY

JANUARY 2006 ◆ VOL 89, NO 1 ◆ SINGLE ISSUE, US $12

TRUTH

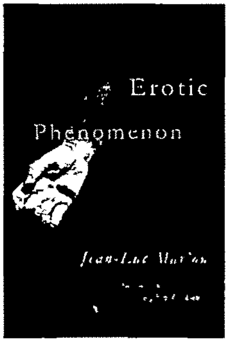
ADVISORY EDITOR JC BEALL

ARTICLES	
JC BEALL	*Introductory Remarks*
HARTRY FIELD	*Compositional Principles* vs *Schematic Reasoning*
ROBERT BARNARD AND TERENCE HORGAN	*Truth as Mediated Correspondence*
FRANK JACKSON	*Representation, Truth, Realism*
MICHAEL P. LYNCH	*ReWrighting Pluralism*
ALEXANDER MILLER	*Russell, Multiple Relations, and the Correspondence Theory of Truth*
NIKOLAJ J. L PEDERSEN	*What Can the Problem of Mixed Inferences Teach Us About Alethic Pluralism?*
ROY T COOK	*There* Are *Non-circular Paradoxes (But Yablo's Isn't One of Them!)*
DOUGLAS EDEN PATTERSON	*Tarski, the Liar, and Inconsistent Languages*
NICHOLAS J J. SMITH	*Semantic Regularity and the Liar Paradox*

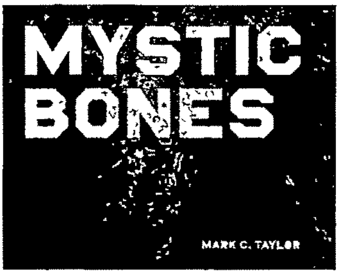
Editor Barry Smith, University at Buffalo State University of New York

Single issues, $12 postpaid Annual subscription institutions US $50, individuals US $30, make checks payable to THE MONIST, 315 Fifth Street, Peru, IL U S A 61354 E-mail philomon1@netscape net

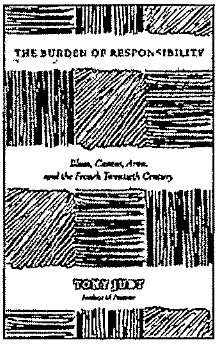
PHILOSOPHY
from Chicago



THE EROTIC PHENOMENON
Jean-Luc Marion
Translated by Stephen E Lewis
"A superb philosophical reflection on love Here, finally, is a book that explores the intimate landscape of each one of us [Marion] deserves our complete attention Let's give it to him " —*L'Express*
CLOTH $35 00



MYSTIC BONES
Mark C. Taylor
"Considering Taylor's moving photographic and written meditation on bones through art, philosophy, and religion, we may divine our own understanding of the creative space between death and life "—Michael Govan, Director, Los Angeles County Museum of Art
CLOTH $25 00


THE BURDEN OF RESPONSIBILITY
Blum, Camus, Aron, and the French Twentieth Century
TONY JUDT

Now in Paperback
THE BURDEN OF RESPONSIBILITY
Blum, Camus, Aron, and the French Twentieth Century
Tony Judt
"An elegantly difficult, eloquent and instructive study of intellectual courage in the face of what the author persuasively describes as intellectual irresponsibility" —Richard Bernstein, *New York Times*
PAPER $16 00

EROTIKON
Essays on Eros, Ancient and Modern
Edited by Shadi Bartsch and Thomas Bartscherer
"*Erotikon* offers a very interesting collection of high quality essays on a subject of more than usual interest "—Peter Toohey, *Bryn Mawr Classical Review*
PAPER $22 50

THE UNIVERSITY OF CHICAGO PRESS • www press uchicago edu

WHAT IS GOOD AND WHY

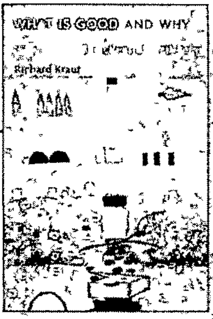
THE ETHICS OF WELL-BEING

RICHARD KRAUT

"Have Rawls and Nozick met their match? The titans of late-twentieth-century social philosophy do indeed find an acute critic—and possible successor—in Kraut."

—Bryce Christensen, *Booklist* (starred review)

new in cloth

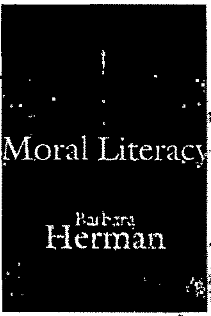


MORAL LITERACY

BARBARA HERMAN

Challenging various Kantian orthodoxies, Herman offers a view of moral competency as a complex achievement, governed by rational norms and dependent on supportive social conditions.

new in cloth



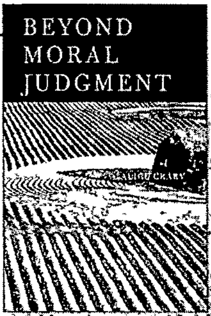
BEYOND MORAL JUDGMENT

ALICE CRARY

"A work of importance in philosophy enables us to bring into view assumptions we had been making, and thus to rethink our questions. This is what *Beyond Moral Judgment* does."

—Cora Diamond

new in cloth



LECTURES ON THE HISTORY OF POLITICAL PHILOSOPHY

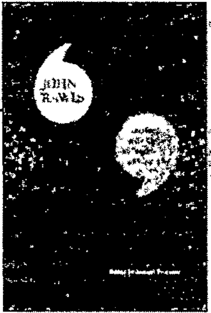
JOHN RAWLS

Edited by Samuel Freeman

"[T]he issuing of this posthumous volume, carefully edited by [Samuel] Freeman, a former student and teaching assistant from Rawls's courses at Harvard University, is a major event."

—David Gordon, *Library Journal*

Belknap Press • new in paperback



WWW.HUP.HARVARD.EDU HARVARD UNIVERSITY PRESS

IN THE SPACE OF REASONS

SELECTED ESSAYS OF WILFRID SELLARS

EDITED BY KEVIN SCHARP AND ROBERT B. BRANDOM

"Sellars was as original a mind as C. S. Peirce, and it has taken almost as long for the importance of his ideas to be appreciated."

—Richard Rorty

new in cloth

MIND IN LIFE

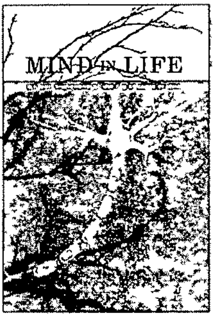
BIOLOGY, PHENOMENOLOGY, AND THE SCIENCES OF MIND

EVAN THOMPSON

"Evan Thompson has emerged as a major presence in the science of the mind. His new book is quite wonderful to read, and I found it impossible to put down."

—Merlin Donald, Case Western Reserve University

Belknap Press • new in cloth



SELF-CONSCIOUSNESS

SEBASTIAN RÖDL

"Dr. Rödl has written an original work in philosophy, a book that is austerely 'analytic' in all the good senses of that term."

—Terry Pinkard, Georgetown University

new in cloth

RE-ENGINEERING PHILOSOPHY FOR LIMITED BEINGS

PIECEWISE APPROXIMATIONS TO REALITY

WILLIAM C. WIMSATT

"Bill Wimsatt is a visionary. He was, and remains, a man ahead of his time."

—Robert Batterman, University of Western Ontario

new in cloth

WWW.HUP.HARVARD.EDU HARVARD UNIVERSITY PRESS

ANNOUNCING:

THE JOURNAL OF PHILOSOPHY

ONLINE

The Journal of Philosophy has added
a print and online subscription option for institutions

SUBSCRIPTION OPTIONS

Libraries and Institutions (print and online)	$180 00
Libraries and Institutions (print only)	$100 00
Individuals	$45 00
Students, retired/unemployed philosophers	$20 00
Postage outside the U S	$15 00

Please consult our website at www journalofphilosophy org
for more details

University
29.10.07

THE JOURNAL OF PHILOSOPHY

VOLUME CIV, NUMBER 4
APRIL 2007

page
163 *Curing Cognitive Hiccups* Andy Clark
A Defense of the Extended Mind
193 *Understanding the New Reductionism* Carl Gillett
The Metaphysics of Science and Compositional Reduction

***Published by* The Journal of Philosophy, Inc.**

THE JOURNAL OF PHILOSOPHY

FOUNDED BY FREDERICK J E WOODBRIDGE AND WENDELL T BUSH

Purpose To publish philosophical articles of current interest and encourage the interchange of ideas, especially the exploration of the borderline between philosophy and other disciplines

Editors Bernard Berofsky, Akeel Bilgrami, John Collins, Arthur C Danto, Kent Greenawalt, Patricia Kitcher, Philip Kitcher, Isaac Levi, Wolfgang Mann, Mary Mothersill, Frederick Neuhouser, Christopher Peacocke, Carol Rovane, Achille C Varzi *Consulting Editors* David Albert, James T Higginbotham, Robert May, Charles D Parsons, Wilfried Sieg *Managing Editor* John Smylie

THE JOURNAL OF PHILOSOPHY is owned and published by the Journal of Philosophy, Inc *President*, Arthur C Danto, *Vice President*, Akeel Bilgrami, *Secretary*, Daniel Shapiro, *Treasurer*, Barbara Gimbel, *Other Trustees* Lee Bollinger, Leigh S Cauman, Kent Greenawalt, Michael J Mooney, Lynn Nesbit

All communications to the Editors and Trustees and all manuscripts may be sent to John Smylie, Managing Editor, Mail Code 4972, 1150 Amsterdam Avenue, Columbia University, New York, New York 10027 FAX (212) 932-3721

You may also visit our website at www journalofphilosophy org

THE JOURNAL OF PHILOSOPHY

2007

SUBSCRIPTIONS (12 issues)

Individuals	$45 00
Libraries and Institutions	$100 00
Libraries and Institutions (print and online)	$180 00
Students, retired/unemployed philosophers	$20 00
Postage outside the U S	$15 00

Payments only in U S currency on a U S bank All back volumes and separate issues available back to 1904 Please inquire for advertising rates, ad space is limited, so ad reservations are required

Published monthly as of January 1977, typeset by Dartmouth Journal Services, Waterbury, VT, and printed by The Sheridan Press, Hanover, PA

All communication about subscriptions and advertisements may be sent to Pamela Ward, Business Manager, Mail Code 4972, 1150 Amsterdam Avenue, Columbia University, New York, NY 10027 (212) 866-1742

All materials published in *The Journal of Philosophy* are copyrighted by the *Journal* and are protected by U S copyright law All rights reserved For more information about our copyright policies, please consult our January issue or contact us

The Journal of Philosophy (ISSN 0022-362X) is published monthly by The Journal of Philosophy, Inc , MC4972, Columbia University, 1150 Amsterdam Avenue, New York, NY 10027 Periodicals postage paid at New York, NY, and other mailing offices

POSTMASTER Please send address changes to *The Journal of Philosophy* at MC4972, Columbia University, 1150 Amsterdam Avenue, New York, NY 10027

© Copyright 2007 by the Journal of Philosophy, Inc ISSN 0022–362X

University

THE JOURNAL OF PHILOSOPHY
VOLUME CIV, NO 4, APRIL 2007

CURING COGNITIVE HICCUPS A DEFENSE OF THE EXTENDED MIND*

Human cognitive processing, according to the *Extended Mind Hypothesis*[1] may at times extend into the environment surrounding the organism[2] Such a view should be contrasted with a nearby, (but much more conservative) view according to which certain cognitive processes lean heavily on environmental structures and scaffoldings, but do not thereby include those structures and scaffoldings themselves This more conservative view may be claimed to capture all that can be of philosophical or scientific interest in such cases, and to avoid some significant dangers into the bargain I shall argue, by contrast, that (in the relevant cases) it is the conservative view that threatens to obscure much that is of value, and that a robust notion of cognitive extension thus earns its keep as part of the emerging picture of the active embodied mind To make this case I first sketch some quite general responses to the worries that motivate the more conservative view I then present some new examples

---

* Thanks to Robert Rupert, Robert Wilson, Susan Hurley, Michael Wheeler, Richard Menary, Mark Rowlands, Dan Hutto, and all who participated in the "Extended Mind 2" conference at the University of Hertfordshire in 2006 for their comments and suggestions on an earlier draft of the present paper

[1] This is introduced in Andy Clark and David Chalmers, "The Extended Mind," *Analysis*, LVIII, 1 (1998) 7–19

[2] Other treatments that encompass notions of cognitive extension include Robert Wilson, *Boundaries of the Mind The Individual in the Fragile Sciences—Cognition* (New York Cambridge, 2004), John Haugeland, "Mind Embodied and Embedded," in Haugeland, *Having Thought Essays in the Metaphysics of Mind* (Cambridge Harvard, 1998), pp 207–40, Daniel Dennett, *Kinds of Minds* (New York Basic Books, 1996), Susan Hurley, *Consciousness in Action* (Cambridge Harvard, 1998), Mark Rowlands, *Externalism Putting Mind and World Back Together Again* (Montreal Acumen/McGill-Queen's, 2003), Michael Wheeler, *Reconstructing the Cognitive World* (Cambridge MIT, 2005), and Richard Menary, *Cognitive Integration Attacking the Bounds of Cognition* (New York Palgrave Macmillan, 2007)

0022-362X/07/0404/163–92

and arguments that aim to flesh out the skeleton responses and to illuminate further the nature and importance of cognitive extension itself

I SPREADING THE MIND

As neuroimaging studies reveal more and more about the role of the brain in a wide variety of cognitive acts and operations, the temptation grows to locate *all* the physical machinery of mind firmly in the head and central nervous system Yet we should not conclude, simply from the undoubted importance of such neural unfoldings in all our cognitive endeavors, that the neural machine and the cognitive machine are always and everywhere simply one and the same Such views seem challenged, for example, by emerging work in "embodied cognitive science"[3] that depicts the nonneural body and the nonbodily world as each capable of making key cognitive contributions According to the more radical versions of these views,[4] the actual local operations that realize certain forms of human cognizing include inextricable tangles of feedback, feed-forward, and feed-around loops loops that promiscuously criss-cross the boundaries of brain, body, and world The local mechanisms of mind, if this is correct, are not all in the head Cognition leaks out into body and world

That may sound like a strange idea But it is hardly stranger, on reflection, than the commonplace idea that the activity of *brain-meat* realizes all that matters about human cognition For to embrace the extended mind is in no way to question the basic materialist vision of mind as emerging fully and without residue from physical goings-on Any added strangeness flows merely from the fact that some of the relevant goings-on, if an extended mind story is correct, do not stay neatly in the brain They do not even stay neatly within the biological body On the contrary, they prove perfectly (and productively) able to span brain, body, and world

---

[3] A representative sample might include Esther Thelen and Linda Smith, *A Dynamic Systems Approach to the Development of Cognition and Action* (Cambridge MIT, 1994), Thelen, G Schöner, C Scheier, and Smith, "The Dynamics of Embodiment A Field Theory of Infant Perseverative Reaching," *Behavioral and Brain Sciences*, xxiv (2001) 1–86, Hurley, *Consciousness in Action*, Andy Clark, *Being There Putting Brain, Body and World Together Again* (Cambridge MIT, 1997), Rowlands, *Body Language Representing in Action* (Cambridge MIT, 2006), Wilson, *Boundaries of the Mind*, Michael Wheeler, *Reconstructing the Cognitive World* (Cambridge MIT, 2005), Alva Noë, *Action in Perception* (Cambridge MIT, 2004), Lawrence Shapiro, *The Mind Incarnate* (Cambridge MIT, 2004), Rolf Pfeifer and Josh Bongard, *How the Body Shapes the Way We Think* (Cambridge MIT, 2007)

[4] For clear examples of the radical tendency (sometimes called "vehicle externalism"), see Hurley, Rowlands, Wilson, and Clark and Chalmers

But might there be compelling reasons to delimit the realm of the *truly cognitive* in the way more standard approaches suggest? Robert Rupert[5] makes just such a claim He distinguishes two projects, which he sees as competing proposals for understanding the cognitive processing of embodied and environmentally situated agents The first (the Extended Mind approach) depicts human cognitive processing as sometimes quite literally extending into the extra-organismic environment Rupert dubs this the *Hypothesis of Extended Cognition* (HEC) and glosses it like this

> (HEC)
> According to this view human cognitive processing literally extends into the environment surrounding the organism, and human cognitive states literally comprise—as wholes do their proper parts—elements in that environment (*ibid*, p 389)

Rupert depicts HEC as a radical hypothesis apt (if true) to transform cognitive scientific theory and practice and to impact our conceptions of agency and persons But it needs to be assessed, Rupert argues, alongside that more conservative (though still interesting and important) competitor This is the perspective he dubs *Hypothesis of Embedded Cognition* (HEMC) according to which

> (HEMC)
> Cognitive processes depend very heavily, in hitherto unexpected ways, on organismically external props and devices and on the structure of the external environment in which cognition takes place (*ibid*, p 393)

The argument starts from a simple appeal to common sense Common sense, Rupert suggests, rebels at the vision of extended cognition[6] so we need sound theoretical reasons to endorse it HEMC, by contrast, is said to be much more compatible with common sense Two main worries are then raised for HEC

The first worry, similar to one previously raised by other critics such as Fred Adams and Kenneth Aizawa,[7] concerns the profound differences that appear to distinguish the inner and outer contributions Thus Rupert suggests that "the external portions of extended

---

[5] Rupert, "Challenges to the Hypothesis of Extended Cognition," this JOURNAL, CI, 8 (August 2004) 389–428

[6] Though I do not take issue with this claim here, it is by no means obvious that common sense itself is committed either to in-the-head or in-the-organism cognition

[7] See Adams and Aizawa, "The Bounds of Cognition," *Philosophical Psychology*, XIV, 1 (2001) 43–64 For some responses, see Clark, "Intrinsic Content, Active Memory, and the Extended Mind," *Analysis*, LXV, 1 (January 2005) 1–11, Clark, "Memento's Revenge The Extended Mind, Re-visited," to appear in Menary, ed , *The Extended Mind* (Aldershot, UK Ashgate, forthcoming)

'memory' states (processes) differ so greatly from internal memories (the process of remembering) that they should be treated as distinct kinds" (*op cit*, p 407) Given these differences, there is said to be no immediate pressure to conceive the internal and the external contribution in the same terms But worse still, there is (allegedly) a significant cost

The second worry thus concerns the apparent scientific cost[8] of any wholesale endorsement of HEC For taking all kinds of external props and aids as proper parts of human cognitive processing robs us, he fears, of the standard object of cognitive scientific theorizing, namely, the stable persisting human individual Even in cases of developmental theorizing, where what is at issue is not so much stability as change, Rupert argues, one still needs to find some persisting, though developing, thinking core Failure to do so, Rupert fears, will cause cognitive science and cognitive psychology to lose their grip on the very human subjects they aim to study The price of HEC, if this were so, is nothing less than the loss of much (perhaps all) the progress that cognitive psychology has made thus far The sciences of the mind, it thus seems, simply cannot afford to identify human cognitive processing with the activity of various "short-lived coupled systems" comprising neural, bodily, and worldly[9] elements

II RESPONDING TO THE CHALLENGES

These are important challenges Nonetheless, the worries are misplaced First, none of the arguments for extended cognition turn on, or otherwise require, the *fine-grained* similarity of the inner and outer contributions Second, HEC need not accrue the prohibitive costs that Rupert fears This section displays the basic form of those replies

Concerning the (lack of) similarity of the inner and outer contributions, part of the problem stems from a persistent misreading of the so-called "Parity Principle" originally introduced by Andy Clark and David Chalmers This was the claim that "if, as we confront some task, a part of the world functions as a process which, were it to go on in the head, we would have no hesitation in accepting as part of the cognitive process, then that part of the world is (for that time) part of the cognitive process "[10] But far from requiring or suggesting the need for any fine-grained similarity between inner and outer processes, the parity claim was specifically meant to *undermine* any tendency to think that the shape of the (present day, human) inner processes sets some bar on what ought to count as part of a genuinely cognitive process The parity probe was thus meant to act as a kind of *veil of metabolic ignorance*, inviting us to ask what our attitude would be if currently external means of storage and transformation were, contrary to the presumed facts, found in biology Thus understood, parity is not about the outer performing just like the (human-specific) inner Rather, it is about equality of opportunity It is about avoiding a rush to judgment based on spatial location alone The parity principle was thus meant to engage our rough sense (based on something like coarse functional role) of what we might intuitively judge to belong to the domain of cognition—rather than, say, that of digestion—but to do so without the pervasive distractions of skin and skull

---

[8] This worry appears briefly in Rupert and is pursued by Rupert at greater length in "Innateness and the Situated Mind," to appear in Philip Robbins and Murat Aydede, eds , *The Cambridge Handbook of Situated Cognition* (New York Cambridge, forthcoming)

[9] This is a loose though familiar usage "Bodily" here means gross-bodily, that is, extra-neural, while "worldly" means "extra-bodily " Clearly, brains are nonetheless parts of bodies, and bodies elements of the world

[10] See Clark and Chalmers, p 8

This point is nicely recognized by Michael Wheeler[11] who notes that the *wrong* way to assess parity of contribution is

> [to] fix the benchmarks for what it is to count as a proper part of a cognitive system by identifying all the details of the causal contribution made by (say) the brain [then by looking] to see if any external elements meet those benchmarks (*ibid*, ms p 3)

To do things that way, Wheeler argues, is to open the door to the implausible and highly chauvinistic thought that only systems whose fine-grained causal profile fully matches that of the human brain can be cognitive systems at all Yet just because some alien neural system failed to match our own biological memory-systems in various ways (perhaps, to borrow an example from Rupert (*op cit*) they fail to exhibit the so-called "generation effect" during recall) we should surely not *thereby* be forced to count the action of such systems as noncognitive, or as not an instance of memory at all The Parity Principle is thus best seen as a demand that we assess the bio-external contributions with the same kind of unbiased vision that even the staunch internalist ought to bring to bear on an alien neural organization It is wholly misconstrued as a demand for fine-grained sameness of processing and storage Rather, it is a call for sameness of opportunity, such that bio-external elements *might* turn out to be parts of the machinery of cognition *even if* their contributions are unlike (perhaps because they are deeply complementary to) those of the biological brain

---

[11] See Wheeler, "Minds, Things and Materiality," in C Renfrew and L Malafouris, eds , *The Cognitive Life of Things* (Cambridge, UK. McDonald Institute for Archaeological Research, forthcoming)

But even once we lay to rest the mistaken reading of the Parity Principle (as requiring fine-grained identity of causal contribution) there remains an important and closely related question The question turns on the issue of natural or explanatory kinds Thus Rupert (*op cit*) questions the idea, certainly present in Clark and Chalmers's original treatment, that treating the organism-notebook system as the supervenience base for some of Otto's dispositional beliefs was to be recommended on grounds of *explanatory unity and power* Rupert's worry then takes as a premise the idea that a kind is natural if it is adverted to by the laws or explanations of a successful science Bio-memory thus meets the requirement as it falls under the laws and explanatory frameworks of a successful science—cognitive psychology or cognitive science more generally But (the argument continues) "extended memory" does not fit the causal profile of memory as described by this body of successful science and hence should not be subsumed under the heading of "memory" at all

It is by no means obvious, however, that all acceptable forms of unification require that all systemic elements behave according to the same laws Indeed, to assume they must do so is simply to beg the question against any science whose target is a genuinely hybrid system (for example, a part-connectionist, part-classical computational organization) In such cases one may, of course, hope to find *additional* principles governing the larger (hybrid) organization itself At this point it is surely worth remembering that the study of extended cognitive systems is just beginning, and it is no wonder that our best current unified understandings target the inner elements alone that is where science has so far (mostly) been looking, after all Nonetheless, it remains the substantive empirical bet of the extended systems theorist that larger hybrid wholes, comprising biological and non-biological elements, will *also* (and more on this below) prove to be the proper objects of sustained scientific study in their own right

A further reason to resist the easy assimilation of HEC into HEMC concerns the nature of the interactions between the internal and the external resources themselves Such interactions, it is important to notice, may be highly complex, nested, and nonlinear As a result there may, in some cases, be no viable means of understanding the behavior and potential of the extended cognitive ensembles by piecemeal decomposition and additive reassembly To understand the integrated operation of the extended thinking system created, for example, by combining pen, paper, graphics programs, and a trained mathematical brain, it may be quite insufficient to attempt to understand and then combine (!) the properties of pens, graphics programs, paper, and brains. This may be insufficient for just the same kinds of reason that, within neuroscience itself, invite us to study not *just* the various major neural substructures and their capacities, but *also* their complex nonlinear interactions and the larger-scale activities in which they participate. Here the larger explanatory targets are whole processing cycles, running on temporary (they are sometimes dubbed "soft-assembled") coalitions of neural resources, arising in response to some specific problem-solving purpose. Such soft-assembled neural packages involve the temporally evolving, often highly re-entrant, activity of multiple populations of neurons spanning a variety of brain areas. But why then suppose that the soft-assemblies most relevant to human cognitive achievements are always and everywhere bounded by skin and skull? Why should we not recognize, in our peculiarly structured and artifact-rich world, a succession of similarly complex hybrid ensembles spanning brain, body, and world?

What, finally, of the allegedly high scientific *costs* of such an enlarged perspective? Here it is important to see that there is no need, in taking extended cognition seriously, to lose our grip on the more-or-less stable, more-or-less persisting biological bundle that lies at the heart of each episode of cognitive soft-assembly. Occasionally, under strict and rare conditions, we may confront genuine extensions of that more-or-less persisting core cases where the persisting, mobile resource bundle is augmented in a potentially permanent manner. But in most other cases, we confront only temporary medleys of information-processing resources comprising an integrated subset of neural activity and bodily and environmental augmentations. The mere fact that such circuits are temporary, however, does not provide sufficient reason to downgrade their cognitive importance. Many purely internal information-processing ensembles are likewise transient creations, generated on-the-spot[12] in response to the particularities of task and context. As Jerry Fodor once put it, in such cases it is "*unstable instantaneous* connectivity that counts."[13] The resulting soft-wired ensembles, in which information flows and is processed in ways apt to the task at hand, do not cease to be important just because they are transient.

---

[12] As just one example, consider the account (D. C. Van Essen, C. H. Anderson, and B. A. Olshausen, "Dynamic Routing Strategies in Sensory, Motor, and Cognitive Processing," in Christopher Koch and Joel Davis, eds., *Large Scale Neuronal Theories of the Brain* (Cambridge: MIT, 1994), pp. 271–99) according to which many neurons and neuronal populations serve not as direct encodings of knowledge or information, but as (dumb) middle managers routing and trafficking the internal flow of information between and within cortical areas. These "control neurons" serve to open and close channels of activity, and allow for the creation of a kind of instantaneous, context-sensitive modular cortical architecture. Related proposals include A. R. Damasio and H. Damasio's notion of "convergence zones," which are neuronal populations which likewise initiate and co-ordinate activity in multiple neuronal groups; see Damasio and Damasio, "Cortical Systems for Retrieval of Concrete Knowledge: The Convergence Zone Framework," in Koch and Davis, eds., pp. 61–74.

Rupert worries that, by taking seriously the notion of cognitive extension in that special subclass of transient cases where the new organizations span brain, body, and world, we lose our grip on the persisting systems that are meant to be our objects of study. Certainly, much work in cognitive and experimental proceeds by assuming that subjects are "persisting, organismically bound cognitive systems."[14] But the correct response here, it seems to me, is to let a thousand flowers bloom. If our avowed goal is to discover the properties of the neural apparatus alone, we might want to impede subjects from using their fingers as counting buffers during an experiment. Similarly, if our goal is to understand what the persisting biological organism alone can do, we might want to restrict the use of all nonbiological props and aids. But if our goal is to unravel the mechanically modulated flow of energy and information that allows an identifiable agent (a Tom, Dick, Jane, or Janet) to solve a certain kind of problem, we should not simply assume that every biologically-motivated surface or barrier forms a cognitively relevant barrier, or that it constitutes an important interface[15] from an information-processing perspective. That this can be done while still respecting experimental requirements is shown, for example, by the careful studies reviewed in section III following.

Notice also that we do not *find* individual agents by first finding their cognitive mechanisms. Instead, we find an agent by finding (roughly speaking) a reliable, easily identifiable physical nexus of perception and action, apparently driven by a persisting and modestly integrated body of goals and knowledge. Then (and only then) do we ask, of some particular problem-solving performance displayed by *that very agent*, what (and where) are the mechanisms that make possible that performance? It is at that point that we may (sometimes) be surprised to find that the target performance depends upon a far wider variety of factors and forces[16] than we initially imagined. In so doing we retain a perfectly good grip on the cognitive agents that are our primary objects of study

---

[13] In Fodor, *The Modularity of Mind* (Cambridge: MIT, 1983), p. 118. See also Fodor, *The Mind Doesn't Work That Way* (Cambridge: MIT, 2001).

[14] See Rupert, "Innateness and the Situated Mind," to appear in Robbins and Aydede, eds. (forthcoming).

[15] For more on the questions concerning interfaces, see Haugeland.

[16] The form of reasoning is thus parallel to that which leads Richard Dawkins to describe the web as part of the "extended phenotype" of the spider (see Dawkins,

It may also be helpful to distinguish two possible explanatory goals hereabouts One is to explain the persistingness of specific cognitive agents The other is to display the machinery that underpins an agent's current mental state or that explains some specific cognitive performance Thus visual cortex, as Chalmers notes (personal communication), may be quite irrelevant to my persistence as a subject (I would persist without it) while still being the supervenience base for some of my current mental states and performances What is really at issue, as far as the claims about cognitive extension are concerned, is which bits of the world make true (by serving as the local mechanistic supervenience base for) certain claims about a subject's here-and-now mental states or cognitive processing

III COGNITIVE IMPARTIALITY

Let us assume (what is surely uncontroversial) that the biological brain is, currently at least, the essential core element in all episodes of individual human cognitive activity A question we may then ask is Does the brain "care about" the nature (biological or nonbiological) or location (organism-bound or organism-external) of the processing and storage resources soft-assembled to tackle some cognitive task?

In an important series of experiments, Wayne Gray and his colleagues have shown (in compelling detail) that it is a mistake to privilege any location or any type of operation in the online soft-assembly of a cognitive routine In the first set of such experiments[17] subjects were required to program an on-screen simulation of a VCR control panel The idea was to manipulate the time-costs of accessing the information (concerning channel, start-time, and so on) needed to program the VCR This information was presented in a window beneath the control panel and was either constantly visible by a flick

---

The Extended Phenotype (New York Oxford, 1982)) and that leads J Scott Turner to treat the sound amplifying ("singing") burrows of the Mole Cricket as external physiological organs (see his book, *The Extended Organism The Physiology of Animal-Built Structures* (Cambridge Harvard, 2000)) In each case we start with a working sense of some baseline concept (phenotype, organ) We then notice that stuff that we do not ordinarily treat in those terms is playing the right kind of role to be considered as belonging to that class This is not about the new stuff working just like the old There is no organ much like a burrow, no animal body much like a web Nor does it require equal stability and permanence the web comes and goes in a way the spider-body does not, and singing burrows, unlike inner organs, may be built, destroyed and rebuilt in new locations

[17] See W D Gray and W t - Fu, "Soft Constraints in Interactive Behavior," *Cognitive Science*, xxviii, 3 (2004) 359–82

of the eyes ("free access group") or available only by moving and clicking the mouse to remove an overlaid opaque cover ("gray box group") There was also a "memory test" group (run under both the above conditions such as, free access and gray box) who, unlike the others, had previously memorized all the information required

What the researchers found was that time-costs of information retrieval, measured in milliseconds, appeared to determine the precise mix of resources (bio-memory, motor actions, shifts of attention) recruited to solve the problem That is, the subjects rapidly settled on whatever strategy yielded (at that phase of the programming) the least-cost (measured by time) information retrieval In fact, they did this even when the fastest mix of resources sacrificed perfect knowledge in the world for imperfect[18] knowledge in the head Only when the in-the-world data could be accessed less effortfully (measured by time) than the data stored in biological memory, was it recruited and calls to the external store "built-in" into the dominant strategy

Gray and Wai-Tat Fu present their results as a challenge to the idea that human cognitive strategies actively favor the use of information in the world over information in the head The pendulum may have swung a little too far, they fear, in favor of reliance on external cognitive scaffolding Instead, they argue that their results show that "the time spent retrieving something from memory is weighed the same as time spent in perceptual-motor activity" and that it is therefore a mistake to "presume the privileged status of any location or type of operation" (*ibid*, pp 378, 380) They thus argue for a level playing field with time-costs of access playing the key role in determining the mix of resources recruited as part of some cognitive routine That is

> The cognitive control of interactive behavior minimizes effort by using a least-effort [measured by time] combination of all the mechanisms available to it All mechanisms or sub-systems are on the table There is no reason to think that one mechanism or subsystem has a privileged status in relation to another(*ibid*, p 380)

Or as Gray and V D. Veksler[19] subsequently express the idea

> The central controller makes no functional distinction between knowledge in-the-head versus in-the-world or the means of acquiring that

---

[18] This is so in the case of all the non-memory-test subjects, who had to pick up the information piecemeal during the course of the experiment

[19] Gray and Veksler, "The Acquisition and Asymmetric Transfer of Interactive Routines," in B G Bara, L Barsalou, and M Bucciarelli, eds , *27th Annual Meeting of the Cognitive Science Society, CogSci2005* (Cognitive Science Society, 2005), pp 809–14

> information (such as eye movement, mouse movement and click, or retrieval from memory) (*op cit*, p 809)

This model is described as a "soft constraints" account of interactive behavior Temporal cost-benefit tradeoffs are said to provide a soft constraint (one that may always be over-ridden by various forms of explicit control) on the precise mix of motoric, perceptual, and bio-memory based resources that will, other things being equal, be automatically recruited to perform a given information-processing task on a given occasion In subsequent work, Gray and colleagues[20] directly compare this to a Minimal Memory Model[21] according to which the resource recruitment process aims to *minimize* the use of bio-memory and to *maximize* the use of environmental support They thus agree (with D H Ballard (*op cit*) and others) that the "embodiment level"[22] (the level at which we observe delicate, short time-scale interactions between motoric, perceptual, and bio-memory based resources) is crucial for much of our problem-solving activity, but differ in their account of exactly how the trade-offs are calculated [23] Where Ballard and others predict a bias towards the use of external encoding and storage, Gray and colleagues depict a level playing field[24] with fine temporal considerations calling the tune "Milliseconds matter, and they matter the same regardless of the type of activity with which they are filled "[25]

For current purposes, the resolution of this particular dispute is less important than the fact of its existence, and the susceptibility of the issue to systematic empirical investigation For we have here a sequence of controlled experiments targeting genuinely hybrid ensembles: soft-assembled coalitions comprising bio-storage, motoric, and perceptual modes of access, and bio-external storage. The work by Gray and colleagues thus provides a clear demonstration of the susceptibility (despite the fears of Rupert, of Adams and Aizawa, and others) of such organizations to quite standard forms of cognitive scientific investigation. Even if Rupert and others are right that key terms such as memory cannot, once extended to the nonbiological domain, serve to pick out fully explanatorily unified kinds, this does not mean that the extended organizations in which they participate are not proper objects of scientific inquiry, emerging and dissolving according to determinable principles, and operating in ways that maximize certain properties and features (in this case, speed of access). To the worry that there will be no unified science of heterogeneously constituted systems, we should reply that there not only can be, but already is, a nascent science both of the recruitment (of sets of neural and extra-neural resources) and of the fine-tuned unfolding of activity in just such heterogeneous ensembles.

---

[20] Gray, C R Sims, Fu, and M J Schoelles, "The Soft Constraints Hypothesis A Rational Analysis Approach to Resource Allocation for Interactive Behavior," *Psychological Review*, cxiii, 3 (2006) 461–82

[21] See D H Ballard, M M Hayhoe, and J B Pelz, "Memory Representations in Natural Tasks," *Journal of Cognitive Neuroscience*, vii, 1 (1995) 66–80 See also Hayhoe, "Vision Using Routines A Functional Account of Vision," *Visual Cognition*, vii (2000) 1–3, 43–64

[22] See Ballard et al , and see Ballard, Hayhoe, P Pook, and R. Rao, "Deictic Codes for the Embodiment of Cognition," *Behavioral and Brain Sciences*, xx (1997) 723–67

[23] This emerging empirical debate between the minimal memory account and the purely time-cost based alternative is evidence, it seems to me, of an important turning point in the study of the embodied mind For in place of a loose coalition of ideas concerning the cognitive importance of body and world, we now begin to see the first stirrings of a science, complete with nuanced disagreements open to empirical investigation by broadly sympathetic practitioners

[24] Gray and Veksler depict the level playing field as populated by a set of possible 'interactive routines' where these are "a complex mixture of elementary cognitive, perceptual, and action operations [that] represent basic patterns of interactive behavior" (p 809)

[25] Gray and Fu, p 364

Gray and colleagues sum up their own preferred model with two claims. The first is that "the control system is indifferent to information source".[26] The second (*ibid.*, p. 478) is that the only bias imposed by biology is that of finding the most cost effective mix of elements available. These very broad conclusions are compatible with a wide variety of possible cost-functions (time taken may not always be the prime or sole determinant). But whatever the cost-function or functions (which may turn out to vary with context and goals) what matters most, for our purposes, is the underlying vision of (what I shall dub) *cognitive impartiality*:

> (Hypothesis of Cognitive Impartiality)
> Our problem-solving performances take shape according to some cost-function or functions that, in the typical course of events, accords no special status or privilege to specific types of operation (motoric, perceptual, introspective) or modes of encoding (in-the-head or in-the-world).

This is, in many ways, a quite natural accompaniment to the Parity Principle itself. It states that the biological control system does not "care about" differences of location or type of resource, but simply uses whatever it can, relative to some cost-benefit trade-off, to get the job done.

---

[26] Gray, Sims, Fu, and Schoelles, p. 478.

IV A BRAIN TEASER

Simple as it may sound, the hypothesis of Cognitive Impartiality hides something of a puzzle, at least for those who would depict cognition as (not just embodied but also) extended For it threatens, unless delicately handled, to undermine the image of cognitive extension[27] in quite a delicious fashion Thus suppose we now ask "Just *what is it* that is so potently impartial concerning its sources of order and information?" The answer looks to be "the biological brain " So *have we not ended up firmly privileging the biological brain in the very act of affirming its own impartiality?*

To see past this worry, we must first notice that there are at least two explanatory targets in the immediate vicinity The first is the recruitment of the extended organization itself We may ask just how, and according to what principles, the various elements (perhaps some subset of neural operations, "deictic" uses of eye movements, gestures (see below), and scribblings) came to combine into a specific soft-assembled information processing device In this process of soft-assembly, the brain surely plays a very special role The second concerns the flow of information and processing in the new soft-assembled extended device Relative to that device we may ask just how information flows and is processed in ways that (hopefully) solve some problem HEC helps us to see that, as far as the second of these explanatory projects goes, the bounds of skin and skull are functionally transparent HEMC, by contrast, both threatens to obscure the scientifically important distinction between the two projects and erects a firm skin-based boundary where the process of recruitment and use marks no boundary at all

The puzzle concerning cognitive impartiality is thus resolved Concerning the process of recruitment, it is indeed the biological brain (or perhaps some of its sub-systems) that is in the driving seat That is to say, it is indeed some neurally-based process of recruitment that (following Gray and colleagues) turns out to be so pointedly unbiased regarding the use of inner versus outer circuits, storage, and operations But once such an organization is in place,[28] it is the flow and transformation of information in (what is often) an extended, distributed system that provides the actual machinery of on-going thought and reason

What this suggests is that in rejecting the vision of human cognitive processing as *organism-bound*, we should not feel forced to deny that it is (in most, perhaps all, real-world cases) *organism-centered* It is indeed primarily (though not solely) the biological organism that, courtesy especially of its potent neural apparatus, spins and maintains (or more minimally selects and exploits) the webs of additional structure that then form parts of the machinery that accomplishes its own cognizing [29] Just as it is the spider-body that spins and maintains the web that then (following Richard Dawkins, *op cit*) constitutes part of its own extended phenotype, so it is the biological human organism that spins, selects, or maintains the webs of cognitive scaffolding that participate in the extended machinery of its own thought and reason [30] Individual cognizing, then, is *organism-centered even if it is not organism-bound*

---

[27] Thanks to Paul Schweizer for pointing this out

[28] In most real-world settings, of course, these two stages (though logically distinct) may be fully or partially temporally overlapping

V THOUGHTFUL GESTURES

At this stage of our discussion, it will help to consider a worked example of extended cognizing in action The example I shall take concerns the role of bodily gesture in thought and reason The case is apt since gesture, though clearly itself an organismic activity, is not merely a *neural* activity Moreover, bodily gesture turns out to exhibit some key features whose applicability looks to outrun the bounds of the organism itself

After an extensive enquiry into the nature and organization of human gesture, the psychologist Susan Goldin-Meadow[31] asks an intriguing question Is gesture simply about the expression of fully-formed thoughts, and thus mainly a prop for inter-agent communication (listeners appreciating meanings through others' gestures) or *might gesture function as part of the actual process of thinking?* Some clues (*ibid*, pp 136–49) that it might be more than merely expressive include

that we do it when talking on the telephone
that we do it when talking to ourselves
that we do it in the dark when no one can see
that gesturing increases with task difficulty
that gesturing increases when speakers must choose between options
that gesturing increases when reasoning about a problem rather than merely describing the problem or a known solution

---

[29] This is not to deny, of course, that much of the spinning is done by social groups of organisms spread out over long swathes of history

[30] One difference is that in the case of the webs of cognitive scaffolding, it is often the human organism acting in concert with existing webs of scaffolding that spins, selects, or maintains new layers of scaffolding, resulting in the powerful process that Kim Sterelny dubs "incremental downstream epistemic engineering" see Sterelny, *Thought in a Hostile World The Evolution of Human Cognition* (Malden, MA Blackwell, 2003)

[31] Goldin-Meadow, *Hearing Gesture How Our Hands Help Us Think* (Cambridge Harvard, 2003)

A deflationist might suggest that most of these effects are easily explained by mere association that gesturing without a viewer is just a habit, installed by our experience of gesturing in the normal communicative context It turns out, however (*ibid*, pp 141–44) that speakers blind from birth, who have never spoken to a visible listener, and never seen others moving their hands as they speak, gesture when they speak Moreover, they do so even when speaking to others they know to be blind [32]

Supposing (for the sake of argument) that gesture does play some kind of active causal role in thinking, just what role might that be? One way to find out is to see what happens when actual physical gesture is removed from the mix of available resources To explore the impact of restricting gesture on thought, Goldin-Meadow and colleagues[33] asked two matched groups of children to memorize a list and then to carry out some mathematical problem-solving before trying to recall the list One group (call it the "free-gesture group") could freely gesture during the intervening math task, the other (call it the "no-gesture group") was told not to gesture The results were that restricting the use of gesture during the intervening mathematical task had a robust and significant detrimental effect on the separate memory task (remembering the list of words) The best explanation, according to Goldin-Meadow, is that the act of gesturing somehow shifts or lightens aspects of the overall neural cognitive load, thus freeing up resources for the memory task

Before pursuing this idea, it is necessary to rule out a rather obvious alternative account According to this alternative account, the effort of remembering *not* to gesture (in the no-gesture group) is *adding* to the load, rather than gesture (in the free gesture group) lessening the load If this were so, the no-gesture group would indeed perform less well, but not because gesturing lightens the load Rather, remembering *not* to gesture increases it As luck would have it, some children and adults spontaneously chose not to gesture during some of the episodes of mathematical problem-solving This allowed the experimenters to compare the effects of removing gesture by instruction and by spontaneous (hence presumably effortless) inclination Memory for the initial task turned out to be equally impaired even when the lack of gesture was a spontaneous choice supporting the claim that the gestures themselves play some active cognitive role [34]

---

[32] See also J Iverson and Goldin-Meadow, "Why People Gesture When They Speak," *Nature*, cccxcvi (1998) 228, and Iverson and Goldin-Meadow, "The Resilience of Gesture in Talk," *Developmental Science*, iv (2001) 416–22

[33] Goldin-Meadow, *Hearing Gesture*, chapter 11, and Goldin-Meadow, H Nusbaum, S Kelly, and S Wagner, "Explaining Math Gesturing Lightens the Load," *Psychological Science*, xii (2001) 516–22

An important hint as to the nature of this active role emerges, Goldin-Meadow argues, when we look at cases of gesture-speech mismatches [35] These are cases where what you say and what you gesture[36] are in conflict, for example, you gesture a one-one mapping while failing to appreciate the importance of such a mapping in your simultaneous vocal attempts at solving the problem Many such cases were found and (importantly) the gestures tended to prefigure the child's consciously finding the right solution in speech at a very slightly later point Even if the right solution was not shortly found, the presence of the apt gesture turned out to be predictive of that child's being able to learn the right solution more easily than others whose gestures showed no such tacit or nascent appreciation

After much conjecture and experiment, Goldin-Meadow is led to the following story (drawing also, as she clearly notes, on the groundbreaking work of David McNeill (*op cit*)) The physical act of gesturing, Goldin-Meadow suggests, plays an active (not merely expressive) role in learning, reasoning, and cognitive change by providing an alternative (analog, motoric, visuo-spatial) representational format In this way

> Gesture expands the set of representational tools available to speakers and listeners It can redundantly reflect information represented through verbal formats or it can augment that information, adding nuances possible only though visual or motor formats [37]

Encodings in that visuo-motor format enter, it is argued, into a kind of ongoing coupled dialectic with encodings in the verbal format Gesture thus continuously informs and alters verbal thinking which is continuously informed and altered by gesture, that is, the two form a genuinely coupled system This coupled dialectic creates points of instability (conflict) whose attempted resolutions move forward our thinking, often (though of course not always) in productive ways The upshot is, according to the developmental psychologists Iverson and

---

[34] See Goldin-Meadow, *Hearing Gesture*, chapter 11, pp 155f There is also rumored to be some older work in which children were simply told to sit on their hands, thus effectively removing the gestural option without adding to the memory load!

[35] Goldin-Meadow, *Hearing Gesture*, chapter 12

[36] Much of Goldin-Meadow is devoted to the task of systematically attributing meaning to spontaneous free gestures See also David McNeill, *Hand and Mind* (Chicago University Press, 1992), and McNeill, *Gesture and Thought* (Chicago University Press, 2005)

[37] Goldin-Meadow, *Hearing Gesture*, p 186

Thelen, is "a dynamic mutuality such that activity in any one component of the system can potentially entrain activity in any other "[38]

But is it (at least partly) the actual physical gestures that matter here, or do they merely reflect the transfer of load between two different neural stores? Does gesturing simply shift the burden from a neural verbal store to a neural visuo-spatial store? If so, then it should be harder to perform a separate spatial memory task when freely gesturing than when not This was tested[39] by replacing the original word-recall task with a spatial one that of recalling the location of dots on a grid The results were unambiguous The availability of gesture still helps (still yields improved performance on the memory task) even when the second task is itself a spatial one

The act of gesturing, all this suggests, is not simply a motor act expressive of some fully neurally-realized process of thought Instead, the physical act of gesturing is part and parcel of a coupled neural-bodily unfolding that is itself usefully seen as an organismically-extended process of thought In gesture we plausibly confront a cognitive process whose implementation involves machinery that loops out beyond the purely neural realm This kind of cognitively pregnant unfolding need not stop at the boundary of the biological organism Something very similar may, as frequently remarked, occur when we are busy writing and thinking at the same time It is not always that fully formed thoughts get committed to paper Rather, the paper provides a medium in which, this time via some kind of coupled neural-scribbling-reading unfolding, we are enabled to explore ways of thinking that might otherwise be unavailable to us Just such a coupled unfolding was eloquently evoked in a famous exchange between Richard Feynman and the historian Charles Weiner

Weiner once remarked casually that [a batch of notes and sketches] represented "a record of [Feynman's] day-to-day work," and Feynman reacted sharply

> "I actually did the work on the paper," he said
> "Well," Weiner said, "the work was done in your head, but the record of it is still here "

---

[38] Iverson and Thelen, "Hand, Mouth and Brain," in Rafael Núñez and Walter J Freeman, eds , *Reclaiming Cognition The Primacy of Action, Intention and Emotion* (Bowling Green, OH Imprint Academic, 1999), pp 19–40, p 37

[39] Goldin-Meadow and Wagner, "How Our Hands Help Us Learn," *Trends in Cognitive Sciences,* IX, 5 (2004) 234–41, Wagner, Nusbaum, and Goldin-Meadow, "Probing the Mental Representation of Gesture Is Handwaving Spatial?" *Journal of Memory and Language,* L (2004) 395–407

> "No, it's not a *record*, not really It's *working* You have to work on paper and this is the paper Okay?"[40]

Feynman is right If, following McNeill and Goldin-Meadow, we allow that actual gestures (not simply their neural pre- or post-cursors) can form part of an individual's cognitive processing, there seems no principled reason to suddenly stop the spread the moment skin meets air [41]

VI MATERIAL CARRIERS

In his most recent work, McNeill[42] offers a clear expression of the view that the physical gestures are actually elements in the cognitive process itself McNeill's work is grounded in extensive empirical case-studies of the use of gesture in free speech The key idea, that McNeill uses to understand and organize these studies, is the notion (briefly mentioned above) of an ongoing imagery-language dialectic in which physical gesture acts as a "material carrier " The phrase "material carrier" is due to L S Vygotsky[43] and is meant to convey the idea of a physical materialization that has systematic cognitive effects But once more, we should not be misled by the image of cognitive effects For according to McNeill, "The concept [of a material carrier] implies that the gesture, *the actual motion of the gesture itself*, is a dimension of thinking "[44] Our free (spontaneous, nonconventional) gestures are not, McNeill argues, merely expressions of or representations of our fully achieved "inner" thoughts but are themselves "thinking in one of its many forms" (*ibid*, p 99) Notice that this is not to say that the gestures do not follow from, and lead to, specific forms of neural activity They do, and McNeill has much to say about the neural systems preferentially involved in the generation and reception of spontaneous gesture (*ibid*, chapters 7 and 8) Rather, it is to see the physical act of gesturing as *part of a unified thought-language-hand system whose coordinated activity has been selected or maintained for its specifically cognitive virtues*

---

[40] The exchange is quoted in James Gleick, *Genius The Life and Science of Richard Feynman* (New York Random House, 1993), p 409 Thanks to Galen Strawson for drawing this material to my attention

[41] Adams and Aizawa depict the neural as the seat of all truly cognitive activity while Rupert (wisely I suspect) prefers the whole organism But in so doing he puts a foot (or so I argue) on a slippery slope For once we allow cognitive processes to become sufficiently hybrid (allowing functional cognitive wholes to be made up of parts as heterogeneous as arm and hand motions plus neural activity) there seems no good reason to stop at the skin

[42] McNeill, *Gesture and Thought*

[43] Vygotsky, *Thought and Language*, Alex Kozulin, trans (Cambridge MIT, 1986)

[44] McNeill, *Gesture and Thought*, p 98

There are important differences between McNeill's account and that of Goldin-Meadow, but they are united in seeing the physical gestures as genuine elements in the cognitive process McNeill stresses the idea of "growth points," described as "the minimal unit of an imagery-language dialectic" (*ibid*, p 105) A growth point is a package of imagistic and linear propositional (linguistic) elements that together form a single idea (for example, both conveying the concept of an antagonistic force as a speaker describes some series of events) The points of productive conflict stressed by Goldin-Meadow are not growth points in this technical sense (see, for example, McNeill, *ibid*, p 137) But they are growth points in another, quite routine sense they are collisions in meaning space, crucially mediated by gestural loops into the physical world, which are able to move our thinking along in productive ways These differences in emphasis do not amount, as far as I can tell, to any deep inconsistency in the underlying models of the cognitive virtues of gesture In each case, the loop into gesture creates a material structure that is available to both speaker and listener And just as that material structure may have a systematic cognitive effect upon the listener, so too it may have a systematic cognitive effect on the speaker Here, McNeill invokes an evolutionary hypothesis that he dubs "Mead's Loop "[45] The background to McNeill's suggestion is the discovery of so-called "mirror neurons " These are neurons, first discovered in the frontal lobes of macaques, that fire both when an animal performs some intentional action and when it sees another animal performing the same action [46] McNeill's suggestion is that our own gestures activate mirror-neuron dominated neural resources so that

> One's own gestures [activate] the part of the brain that responds to intentional actions, including gestures, by someone else, and thus treats one's own gesture as a social stimulus [47]

Whether this is the correct evolutionary and mechanistic account is unimportant for present purposes What matters is rather the guiding idea that by materializing imagistic thought in physical gesture we create a stable physical presence that may productively impact and constrain the neural elements of thought and reason The role of gesture, if this is correct, is closely akin to that of certain forms of self-directed, overt or covert speech or (looping outside the organismic shell) to certain forms of writing-for-thinking (see McNeill, *ibid.*, p 99)

---

[45] After George Herbert Mead See Charles W Morris, ed , *Mind, Self, and Society* (Chicago University Press, 1934)

[46] See, for example, G Rizzolatti, L Fogassi, and V Gallese, "Neurophysiological Mechanisms Underlying the Understanding and Imitation of Action," *Nature Review Neuroscience*, ii (2001) 661–70

[47] McNeill, *Gesture and Thought*, p 250

Shaun Gallagher, in a very rewarding recent discussion of gesture[48] and thought writes that

> Even if we are not explicitly aware of our gestures, and even in circumstances where they contribute nothing to the communicative process, they may contribute implicitly to the shaping of our cognition (*ibid*, p 121)

Gallagher approaches the topic of gesture in the larger framework of his account of the "prenoetic" role of embodiment This is a term of art that Gallagher uses to signify the role of the body in structuring mind and consciousness The idea is that facts about the body, and about bodily orientation, and so on, set the scene for conscious acts of perception, memory, and judgment (the "noetic" factors) in various important ways A prenoetic performance, we are told, is "one that helps to structure consciousness but that does not explicitly show itself in the contents of consciousness" (*ibid*, p 32) Thus, to take a very simple example, embodied agents perceive the world from a certain spatial perspective That perspective shapes what is given to us explicitly in phenomenal experience, but it is not itself part of what we experience Instead, it "shapes" or "structures" experience (for this example, see Gallagher, *ibid*, p 2–3) In this manner, Gallagher speaks of the role of gesture in "shaping" cognition and (following Maurice Merleau-Ponty's usage in describing the cognitive role of speech) in the "accomplishment of thought " Such locutions neatly (though only temporarily) sidestep the thorny issue of whether to see gesture as part of the *actual machinery* of thought and reason In a footnote to the quoted passage Gallagher is less cautious, suggesting that

> It may be that certain aspects of what we call the mind just are in fact nothing other than what we tend to call expression, that is, occurrent linguistic practices ("internal speech"), gesture, and expressive movement (Gallagher, *ibid*, p 121)

Gesture, Gallagher here suspects, is both a means by which thought is "accomplished" *and* an aspect of mind—an aspect of the thinking[49] itself

---

[48] Gallagher, *How the Body Shapes the Mind* (New York Oxford, 2005)

[49] There is, unfortunately, substantial ambiguity in the notion of "thinking" invoked in many of these discussions of gesture, since it can sometimes mean (1) "verbal

What finally emerges from our consideration of the role of gesture in thought is that gestures form part of an integrated "language-thought-hand"[50] system, and that it is activity in this whole interlinked system that has been selected for its specifically cognitive virtues. Neural systems coordinate with, help produce, exploit, and can themselves be entrained by, those special-purpose bodily motions that constitute free gestures. In this way speech, as Jana M. Iverson and Esther Thelen noted (*op. cit.*), gesture and neural activity are able to form a single integrated system with clear problem-solving virtues not reducible to the virtues of any of its individual parts.

VII. LOOPS AS MECHANISMS

A single integrated system can have a variety of distinct parts whose contributions to some overall process are hugely different. Some of those parts, moreover, may be cognitive processes in their own right (that is to say, they remain cognitive processes even when considered in isolation from the others) and others not. Just so, a sequence of physical gestures alone could never implement a cognitive state. It is only in coordination with crucial forms of neural activity that the cognitive role of the gestures can emerge and be maintained. By contrast, some set of neural goings-on is often sufficient for the presence of some cognitive state or other. But this (genuine) asymmetry provides no reason to reject the notion that gestures form part of the machinery of cognition. To see this, we need only remind ourselves that the activity of a single neuron is (likewise) never sufficient for the existence of a cognitive state, yet that activity can, in the proper context, still form part of the machinery that implements a cognitive state or process.

It may (or may not) also be true that for any gesture-involving cognitive unfolding, there is a pure sequence of neural events such that *if* they were somehow held in place, or ushered into being without the loop through physical gesture, the cognitive states of the

---

thought" which is conceived, by Goldin-Meadow, as distinct from (though intertwined with) (2) the kinds of holistic, imagistic thinking specifically accomplished by gesture. Finally, there is (3) the overall cognitive state achieved by an agent who has engaged in some ongoing process involving both gestural and verbal elements. To say that gesturing is part of the process that constitutes thinking is thus to say both that it helps mediate and inform the verbal thinkings, and that in so doing it forms part of a larger integrated cognitive system.

[50] The use of "thought" here is misleading (see previous note). It reflects common usage rather than the actual model than McNeill and others develop. There is a similar ambiguity in the use of "language" since gesturing, on McNeill's account, is actually part of language. McNeill is aware of these infelicities but thinks the usage will do no harm: see "Terminological Tango" on page 21 of McNeill.

embodied agent would be the same. It does not follow from this that the gestures play only a causal role and do not help constitute the machinery of cognition. For the same may also be true of a sequence of neural states held together by some internal operation. Achieve that very sequence some other way and the chain of thoughts, let us assume, will come out the same. It does not follow[51] that the inner or the outer operations involved are thereby not (as things actually unfold) genuine aspects of the cognitive process. Thus Susan Hurley[52] usefully cautions against what she dubs "the 'causal-constitutive error' error" where this is

> the error of objecting that externalist explanations give a constitutive role to external factors that are "merely causal" while assuming without independent argument or criteria that the causal/constitutive distinction coincides with some external/internal boundary. To avoid thus begging the question, we should not operate with prior assumptions about where to place the causal/constitutive boundary, but wait on the results of explanation (*ibid.*, ms. p. 5).

In trying to get a grip on these matters we are easily misled by various inessential features of many common cases where bio-external factors and forces impact thought and reason. Thus suppose the rhythmic pulse of rain on my Edinburgh window somehow helps the pace and sequencing of a flow of thoughts. Is the rain now part of my cognitive engine? No. It is merely the backdrop against which my cognizing takes shape. But this, I submit, is not because the rain is outside the bounds of skin and skull. Rather, it is because the rain is not part of (it is not even a side-effect or a "spandrel" within) any system either selected or maintained for the support of better cognizing. It is indeed *mere* (but as it happens helpful) backdrop. Compare this with a robot *designed* to use raindrop sounds to time and pace certain internal operations essential to some kinds of problem-solving. Such a robot would be vulnerable to (non-British) weather. But it is not clear (at least to me) that the whole drop-based timing mechanism is not usefully considered as one of the robot's cognitive routines. Consider finally the Self-Stimulating Spitting Robot. This is a robot that evolved to spit stored water at a plate on its own body for the same purpose, so as to use the auditory signal as a kind of "virtual wire"[53] to time other key operations Those self-maintained cognition-supporting signals are surely part of the cognitive mechanism itself A neural clock or oscillator would count after all

---

[51] See Hurley, *Consciousness in Action*, chapter 8, for more sophisticated versions of this argument.

[52] Hurley, "The Varieties of Externalism," to appear in Aydede and Robbins, eds. (forthcoming).

[53] This notion can be found in Daniel Dennett's *Consciousness Explained* (Boston: Little, Brown, 1991), p. 196.

What these simple examples show is that (as we might expect) coupling alone is not enough *Sometimes*, all coupling does is provide a channel allowing externally originating inputs to drive cognitive processing along But in a wide range of the most interesting cases there is a crucially important complication These are the cases where we confront a recognizably cognitive process, running in some agent, that creates outputs (speech, gesture, expressive movements, written words) that, recycled as inputs, drive the cognitive process along In such cases, any intuitive ban on counting *inputs* as parts of *mechanisms* seems wrong Instead, we confront something rather like the cognitive equivalent of a forced induction system A familiar example is the turbo-driven automobile engine The turbocharger uses exhaust flow from the engine to spin a turbine that spins an air pump that compresses the air flowing into the engine The compression squeezes more air into each cylinder, allowing more fuel to be combined, leading to more powerful explosions (that drive the engine that creates the exhaust flow that powers the turbo) This self-stimulating automotive arrangement provides up to 40% more power on demand The exhaust flow is an engine output in good standing that also serves as a reliable, self-generated input There can surely be little doubt that the whole turbocharging cycle should count as part of the automobile's own overall power-generating mechanism! The same is true, I submit, in the case of gesture gesture is both a systemic output and a self-generated input[54] that plays an important role in an extended neural-bodily cognitive economy The wrong image here is that of a central reasoning engine that merely uses gesture to clothe or materialize pre-formed ideas Instead, gesture (and overt or covert speech) emerge as interacting parts of a distributed cognitive engine, participating in cognitively potent self-stimulating loops whose activity is as much an *aspect* of our thinking as its *result*

VIII PARTICIPANT MACHINERY AND MORPHOLOGICAL COMPUTATION

The goal of this treatment has been to defend the claim that (non-neural) body and (extra-bodily) world can, and sometimes do, act as what might now be dubbed "participant machinery" Body and world, if this is correct, sometimes form part of the very machinery by which mind and cognition are physically realized, and hence form part of the local material supervenience base for various mental states and processes. Since this has seemed to many quite an exotic claim, it may be worth ending with one more (just about maximally simple, but nonetheless revealing) illustration.

---

[54] In the case of gesture the relation between the self-created inputs and other processing elements also looks to involve the full complexities of "continuous reciprocal causation" as discussed in my *Being There Putting Brain, Body and World Together Again* (Cambridge MIT, 1997), pp 163–66)

Chandana Paul describes a toy example designed to demonstrate "that a robot body can be used for computation in addition to merely acting as an effector for the controller."[55] The backdrop to the demonstration involves a very simple class of neural networks known as perceptrons.[56] It is well known that a perceptron, if given two inputs A and B, can compute OR and AND functions (in fact, all linearly separable functions) but not linearly inseparable ones such as exclusive-or. Exclusive-or, normally written XOR, is true if either *but not both* disjuncts are true: that is, if (A or B) is true but – (A and B). Paul's demonstration involves a simple "vehicle," of the kind made famous by Valentino Braitenberg,[57] whose behavior is determined by the activity of two perceptron networks, each with inputs A and B. Perceptron 1 computes OR and controls M1, a forward drive delivered to the single central front wheel of a front-wheel drive vehicle. This means that power is delivered to the single central front wheel if either or both inputs are active (it is thus computing the standard INCLUSIVE OR function). Perceptron 2 computes the standard AND function and controls M2, a lifting device that will raise the single front wheel of the forward drive vehicle off the ground if and only if *both* inputs are active. You can probably see where this is going. When A and B are both reading OFF (zero, false), both nets output zero, the wheel is on the ground, but no power is delivered so the robot stays stationary. When only A is ON, the AND net delivers zero, the wheel stays grounded, and the OR net outputs a one. The wheel turns and the robot goes forward. The same type of scenario occurs when only B is ON. But (the crucial case) when A and B are both ON the OR net causes M1 to move but the AND net lifts the wheel from the ground so the robot stays stationary. The embodied systems response profile to the different possible values of the A and B inputs thus has the form of the standard XOR truth table, despite the fact that the visible computational controllers are perceptrons, congenitally unable to compute nonlinearly separable functions such as XOR Lifting the front wheel in response to the conjunction of the two inputs now stands in for the "missing line" of the XOR truth table The physical vehicle, despite having only perceptrons for controllers, is thus able to behave exactly as if it were controlled by an XOR net, because the active body of the robot is providing the functional equivalent of the missing second layer of neural processing The overall embodied system thus provides the missing functionality, equivalent to performing a NOT on the first input, followed by an AND In this way, Paul concludes "the example shows that through its configuration a robot body can perform a quantifiable operation on its inputs "[58]

---

[55] Paul, "Morphology and Computation," in Stefan Schaal, Auke Jan Ijspeert, Aude Billard, Sethu Vijayakumar, John Hallam, and Jean-Arcady Meyer, eds., *From Animals to Animats, Proceedings of the 8th International Conference on the Simulation of Adaptive Behaviour, Los Angeles, CA, USA, 2004* (Cambridge: MIT, 2004), pp. 33–38, quoted passage p. 33. An expanded version of this paper appears as "Morphological Computation: A Basis for the Analysis of Morphology and Control Requirements," in *Robotics and Autonomous Systems*, LIV (2006): 619–30.

[56] Frank Rosenblatt, *Principles of Neurodynamics* (New York: Spartan Books, 1962).

[57] Braitenberg, *Vehicles: Experiments in Synthetic Psychology* (Cambridge: MIT, 1984).

At this point a skeptic might argue that the XOR computation is in some way unreal more in the eye of the observer than a true resource for a reasoning robot? And there is (as things stand) some truth in this For what the robot currently displays is what Paul nicely dubs "latent morphological computation" computation that is visibly (to us) implicit in the response profile of the overall physical device, but not yet available *to the device itself* as a general-purpose problem-solving resource

A simple (and, as we shall see, biologically unexceptional) tweak, however, makes the new functionality available to the device itself Thus Paul next describes a "vacuum cleaning robot" (the precise details of which need not concern us here) The vacuum cleaning robot is like the XOR robot except that this time it is augmented with a sensor informing it of the behavioral consequences of its own action Thus augmented, the robot can learn (or be programmed) so as to incorporate the body-involving XOR circuit into an open-ended set of other routines, routing various A, B signals through the body circuit and reading the XOR result off from a rapid, self-perceived bodily "twitch" of the front wheel a twitch that need not even persist long enough to cause actual forward motion The body-involving XOR computation (that may previously have appeared to be merely in the eye of the beholder) is thus scaled up to a general-purpose resource that can be invoked much like a regular logic gate Quite generally then

> When a robot with latent morphological computation is augmented with a sensor which can sense the behavioral consequences, it makes the computational function defined by the morphology explicit, such that it

---

[58] Paul, "Morphology and Computation," p 34

> can be used as a standard computational sub-unit at any stage of the processing (Paul, *op cit*, p 36)

It might seem, nonetheless, that this is still just a rather clumsy trick Why use the robot body to perform a computation that could be so cheaply and easily handled using a simple three layer neural network? This is, however, to miss the true point and force of the demonstration For the idea is that evolved biological intelligences, unlike the more neatly engineered solutions with which we are still most familiar as designers, are perfectly able to find and exploit unexpected forms of *multiple functionality* That is to say, they may find and exploit solutions in which a single element (such as a bodily routine or motion) plays many roles, some of them merely practical, others more "epistemic"[59] in nature The clean division between mechanical (body) design and controller design that characterizes many humanly engineered solutions looks quite unimportant (indeed, often counterproductive) if what we seek is efficiency and maximal exploitation of resources Paul's demonstration may be compared to the work of Adrian Thompson and colleagues[60] using genetic algorithms to artificially evolve real electronic circuits The evolved circuits turned out to exploit all manner of physical properties, usually ignored or deliberately suppressed by human engineers [61] The lesson, according to the authors, was that

> It can be expected that all of the detailed physics of the hardware will be brought to bear on the problem at hand time delays, parasitic capacitances, cross-talk, meta-stability constraints and other low-level characteristics might all be used in generating the evolved behavior (Thompson et al, *op cit*, p 21)

What thus goes for the brain (the hardware chip) goes too for the rest of the physical body It too may be exploited, in all manner of unexpected ways, as an essential part of an information-processing organization In real-world cases, Paul goes on to suggest, we should expect to find that the computational roles being played by bodily acts are much more complex than the computation of a common binary function, perhaps involving analog functions of quite unexpected degrees of complexity The case of gesture-for-thought (section VI above) may be an example of just this kind, in which actual hand and arm motions look to implement encoding and processing operations that are (as McNeill suggests) holistic and analog rather than local, symbolic, and discrete

---

[59] For more on this, see David Kirsh and Paul Maglio, "On Distinguishing Epistemic from Pragmatic Action," *Cognitive Science*, XVIII (1994) 513–49

[60] A Thompson, I Harvey, and P Husbands, "Unconstrained Evolution and Hard Consequences," in Eduardo Sanchez and Marco Tomassini, eds, *Towards Evolvable Hardware* (Boston Springer, 1996), pp 136–65

[61] For further discussion, see chapter 5 of my *Mindware An Introduction to the Philosophy of Cognitive Science* (New York Oxford, 2001)

Once we start to question our simplistic armchair visions of the division of labor between brain, body, and world it rapidly becomes clear that there is no barrier to the realization of key functional and/or computational organizations by complex mixtures of neural, bodily, and environmental structures The simple robotic illustration also provides further evidence of the power and importance of cognitive self-stimulation (section VII above) For the step from latent to explicit morphological computation depends essentially upon the agent's ability to sense its own bodily states As embodied agents replete both with systems for efferent copy and for sensing what our own bodies are doing, we are ideally placed to profit from our own bodily actions, and to exploit our own bodily acts for cognitive and computational ends Daily embodied activity may thus be playing many subtle, yet-to-be-understood cognitive roles [62]

IX WHY THE HEC?

This paper began with a double challenge Show us that HEC (Hypothesis of Extended Cognition) has not priced itself out of the market by depicting us as cognitively extended agents at the cost of identifying persisting subjects for scientific study, and show us that there is real added value in adopting the perspective of HEC rather than its more innocuous-seeming cousin HEMC (Hypothesis of Embedded Cognition)

---

[62] To take just one concrete example, there is a growing body of work on the possible role of eye movements in thought, reason, discourse comprehension, and recall Such work includes Ballard and colleagues's work on deictic (fixation-based) pointers in block-copying (Ballard, Hayhoe, Pook, and Rao (*op cit*)), D C Richardson and M J Spivey's account of the cognitive role of eye movements in recall (Richardson and Spivey, "Representation, Space and Hollywood Squares Looking at Things That Aren't There Anymore," *Cognition*, LXXVI (2000) 269–95), Richardson and N Z Kirkham's exploration of the role of eye movements (in 6 month olds) as spatially indexing auditory information (Richardson and Kirkham, "Multi-modal Events and Moving Locations Eye Movements of Adults and 6-Month-Olds Reveal Dynamic Spatial Indexing," *Journal of Experimental Psychology General*, CXXXI, 1 (2004) 46–66), and Richardson and R Dale's model of the role of coupling between speaker and listener's eye-movements in discourse comprehension (Richardson and Dale, "Looking to Understand The Coupling between Speakers' and Listeners' Eye Movements and Its Relationship to Discourse Comprehension," *Proceedings of the Twenty-sixth Annual Meeting of the Cognitive Science Society* (Mawhah, NJ Erlbaum, 2004), pp 1143–48 For a useful review, see Spivey, Richardson, and S Fitneva, "Thinking Outside the Brain Spatial Indices to Linguistic and Visual Information," in J Henderson and F Ferreira, eds , *The Interface of Vision Language and Action* (New York Psychology Press, forthcoming)

Both challenges have now been met Concerning the first, we have arrived at a vision of human cognition as organism-centered but not organism-bound Embracing HEC does not require us to abandon the vision of a persisting biological (and within the biological, a neural) core that is a perfectly proper object of cognitive scientific study HEC simply asserts that we should also study larger, often temporary, ensembles as units of cognitive activity in their own right

At that point the second challenge becomes pressing What is the added value accruing to the choice of HEC over HEMC? Both sides should, I think, concede the in-principle availability of alternative ways of carving the cognitive cake As with any science, the placement of systemic boundaries reflects both general considerations and specific explanatory ambitions The added value (for some purposes) of the HEC-style parse emerges, or so I have tried to show, once we recognize a few less familiar facts concerning human cognitive organization It emerges when we recognize that the brain/CNS is "cognitively impartial" it does not care how and where key operations are performed And it emerges when we recognize that much human cognizing benefits quite profoundly from cycles of self-stimulating activity ("cognitive turbo-drives") in which we actively create the structures that drive and constrain our own evolving processes of thought and reason

It seems likely that some opposition to the idea of extended cognitive systems is rooted in the imagined availability of a simple alternative model in which a skull-bound intelligent agent decides to offload[63] certain bits of work and storage onto bodily and environmental structures In the cases at issue, however, there is no such act of conscious offloading and reloading to be found We do not, for example, consciously choose to gesture so as to lighten or alter the neuro-cognitive load In such cases the extended process involves complex *sub-personally integrated* routines that are (typically) either selected or maintained for their peculiarly cognitive virtues Nor need we imagine, in cases where conscious choice and orchestration is missing, that some highly intelligent, well-informed, though as-it-happens non-conscious, inner executive has made the choice for us

---

[63] For example, we decide to place that yellow sticky on the mirror reminding us of the important meeting in the morning In such cases an identifiable, fully-contained thinking agent offloads onto some environmental structure the semantically well-formed product of some recognizably cognitive activity, only to later re-load it as needed to perform some task

Instead the "choice" consists simply in the emergence of an effective distributed problem-solving ensemble, where such emergence is guided by principles we are only just beginning (as in the work by Ballard, Gray, and others) to understand. A serious worry about the apparently innocent and conservative HEMC, it seems to me, is thus that it threatens to repeat for outer circuits and elements the mistake that Daniel Dennett (*op. cit.*) famously warns us against with regard to inner circuits and elements. It depicts such bodily or outer resources as doing their work only by parading structure and information in front of some thoughtful neural overseer. In the absence of any such privileged inner component, the outer and the inner operations are free to emerge as well-tuned co-active participants in the construction of thought and reason.

To be sure, this overall vision (of cognition genuinely distributed between brain, body, and world) bequeaths a brand new set of puzzles. It invokes an ill-understood process that soft-assembles a problem-solving whole from a candidate pool that may include neural storage and processing routines, perceptual and motoric routines, external storage and operations, and a variety of self-stimulating cycles involving self-produced material scaffolding. But importantly, this all applies with equal force to the neural economy itself. Here too, a cognitive task will often be addressed by a soft-assembled coalition of distributed (and often highly heterogeneous) neural components and brain areas, temporarily held together by a transient pattern of "functional connectivity."[64] HEC thus gains in plausibility as soon as the *inner* economy is seen aright: as itself hybrid, fragmented, yet vastly empowered by an ill-understood capacity to form and re-form into a variety of surprisingly integrated (though temporary) wholes. What HEC allows us to see clearly is that where ongoing human cognitive activity is concerned, there are usually *many* boundaries in play, *many different kinds* of capacity and resource in action, and a complex flux of recruitment, retrieval, and processing defined across these shifting, heterogeneous, multi-faceted wholes.

This turns Rupert's argument on its head. For having allowed that we could, if we so wished, choose to parse our cognitively potent coupled unfoldings according to either HEC or HEMC, we can now see that it is the choice of HEMC that sometimes threatens to obscure much that is of value. We do indeed seek to carve nature at the most

---

[64] For a nice account of the distinction between persisting structure and short-lived functional complexity in the brain, see O. Sporns, D. Chialvo, M. Kaiser, and C. C. Hilgetag, "Organization, Development and Function of Complex Brain Networks," *Trends in Cognitive Sciences*, VIII (2004): 418–25.

Calcutta University Library ALIPORE CAMPUS

causally relevant joints, a task not accomplished by elevating anatomic and metabolic boundaries into make-or-break cognitive ones The cure for cognitive hiccups (the unproductive argumentative oscillation from HEC to HEMC to HEC to HEMC) is thus at hand For the only real danger from HEC is that it may blind us to the genuine extent to which human cognition, though not organism-bound, remains importantly organism-centered To guard against that misreading, we may now scout

> HOC (Hypothesis of Organism-Centered Cognition)
> Human cognitive processing (sometimes) literally extends into the environment surrounding the organism But the organism (and within the organism, the brain/CNS) remains the core and currently the most active element Cognition is organism-centered even when it is not organism-bound

HEC, HEMC, HOC? We should not feel locked into some pale zero sum game As philosophers and as cognitive scientists we can (and should) practice the art of flipping between these different perspectives, treating each as a lens apt to draw attention to certain features, regularities, and contributions while making it harder to spot others, or to give them their problem-solving due The cure for cognitive hiccups is to stop worrying and enjoy the ride

ANDY CLARK

University of Edinburgh

UNDERSTANDING THE NEW REDUCTIONISM: THE METAPHYSICS OF SCIENCE AND COMPOSITIONAL REDUCTION*

> The philosophical emptiness of Nagel reduction ... can plainly be seen. ... [Thus] the antireductionism based on the unavailability of ... bridge laws—most importantly, the multiple realization argument of Putnam and Fodor—is irrelevant to the real issue of ... reduction. ... Much of the debate over the past two decades about reductionism has been carried on in terms of an inappropriate model of reduction, and now largely appears beside the point for issues of real philosophical significance.—Jaegwon Kim[1]

> Science is the characteristic product of our culture. Similarly, understanding where science fits in ... is our characteristic philosophical problem. ... As of now, the hardest part is to reconcile a physicalistic ontology with the apparently ineliminable multiplicity of discourses that we require when we try to say how things are.—Jerry Fodor[2]

Our opening passage will likely be a little shocking for anyone unfamiliar with recent work at the intersection of the philosophy of mind and philosophy of science, since one might conclude from such passages that Jaegwon Kim, for so long one of its foremost proponents, has abandoned reductionism! Whatever next you may ask? Is Jerry Fodor retraining in brain-imaging techniques? Actually, things have not changed quite so radically, although a seismic shift is slowly impacting a range of debates and issues. For Kim has not renounced reductionism, but simply the "classical" or "Nagelian" approach, and its numerous descendents, with their focus on derivational relations between semantic entities such as law statements.[3] Instead, Kim has championed the necessity of a

---

* For helpful discussion, I am grateful to Ted Morris, Ernan McMullin, Mike Rea, Robert Richardson, Bradley Rives, Mark Siderits, and, especially, Jerry Fodor, Jaegwon Kim, Barry Loewer, and the editors. I am also grateful to the participants in the Tilburg-Nijmegen University 2003 and 2004 philosophy workshops, and work-in-progress groups at the University of California-San Diego and the University of Notre Dame, where I presented the paper.

[1] Kim, "Making Sense of Emergence," *Philosophical Studies*, xcv (1999): 3–44, here pp. 12–13.

[2] Fodor, "Look! Review of *Consilience* by E.O. Wilson," *The London Review of Books*, xx, 21 (1998): 5–6, here p. 6.

[3] See, for example, Kim's *Mind in a Physical World* (Cambridge: MIT, 1998), and especially his "Making Sense of Emergence." The classical approach is laid out in Ernest Nagel, *The Structure of Science* (New York: Harcourt Brace, 1961).

0022-362X/07/0404/193–216

new approach to reduction that is thoroughly metaphysical in character, centered upon the implications of the ontological relations posited in the ubiquitous type of scientific reasoning I will term *mechanistic* explanation.[4]

Such explanations stand at the foundation of the nonreductive physicalism that is the reigning philosophical orthodoxy, for they were famously used by antireductionist writers, like Fodor, Hilary Putnam, William Wimsatt, Philip Kitcher, and others, to overturn classical reductionism given its reliance upon property identities.[5] These writers argued that mechanistic explanations show that in the sciences we usually only have so-called "realization" relations, not identities, where the associated scientific findings were consequently used to support a view of the natural world as a "layered" compositional hierarchy. Against this backdrop, the ingenuity—even audacity—of Kim's strategy becomes plain, since his idea is that the reductionist should embrace the antireductionist's key insights about special sciences in combination with a certain kind of metaphysical approach. For Kim contends that realization is consequently converted from the shield of antireductionism into the engine of ontological reduction.

This new approach reverses a number of longstanding commitments in debates over reduction and in the philosophy of science more generally. Reductionists, and many other philosophers of science, have until very recently implicitly followed the positivists' proscription of metaphysics and the ontological relations it studies. However, at the heart of Kim's strategy is the suggestion that reductionism must embrace metaphysical issues in the sciences where the required "metaphysics" is not the *a priori* inquiry rightly disparaged by the positivists, but is instead what has recently been dubbed the "metaphysics of science": the careful, abstract investigation of ontological issues as they arise *within* the sciences and their findings, models, explanations, and so on.

---

[4] There are obviously both intralevel and interlevel mechanisms and mechanistic explanations, but here the reader should note that I will primarily be focused upon the latter. For accounts of mechanistic explanation, see, among others, Fodor, *Psychological Explanation* (New York: Random House, 1968), William Wimsatt, "Reductive Explanation: A Functional Account," in R. Cohen et al., eds., *PSA Proceedings 1974* (Boston: Reidel, 1974), Robert Cummins, *The Nature of Psychological Explanation* (Cambridge: MIT, 1975), William Bechtel and Robert Richardson, *Discovering Complexity* (Chicago: University Press, 1993), and P. Machamer, L. Darden, and C. Craver, "Thinking about Mechanisms," *Philosophy of Science*, LXVII (2000): 1–25.

[5] See, for example, Putnam, "Psychological Predicates," in W.H. Capitan and D.D. Merrill, eds., *Art, Mind, and Religion* (Pittsburgh: University Press, 1967), pp. 37–48, Fodor, *Psychological Explanation*, and "Special Sciences," *Synthese*, XXVIII (1974): 97–115, Wimsatt, "Reductionism, Levels of Organization and the Mind-Body Problem," in Gordon Globus, Irwin Savodnik, and Grover Maxwell, eds., *Consciousness and the Brain* (New York: Plenum, 1976), pp. 199–267, and Kitcher, "1953 and All That," *Philosophical Review*, XCIII (1984): 335–73.

My goal in this paper is to examine the approach to reduction that results when Kim's pioneering strategy is fully implemented. I begin, in section I, by using an example from the neurosciences to illuminate the key features of the compositional relations between powers, properties, individuals, and mechanisms that are posited in mechanistic explanations. Once they are properly understood, I argue that such compositional notions drive a powerful form of ontological reduction, what I dub *compositional reductionism*, which takes our initial ontological commitments to both component entities and composed higher-level powers, properties, individuals and processes, and reduces them to a commitment solely to *one* layer of truthmakers in the relevant component entities. Having illuminated the foundation of compositional reductionism, I then explore its broader implications, in section II, showing that these are very different from earlier reductionist positions. I argue that the compositional reductionist should accept many of the claims about special sciences that previous reductionists have disputed, for my discussion will ultimately establish that compositional reductionism is best understood as combining a trenchant ontological reductionism with a semantic antireductionism about special sciences and their predicates.

Such a position often initially appears puzzling, even incoherent, to many philosophers of science, but I suggest that it provides an overlooked, and very important, response to our "characteristic philosophical problem" highlighted by Fodor in our other opening passage. For I show that compositional reductionism can endorse the ineliminability of the discourses of special sciences whilst *still* holding a thoroughgoing ontological reductionism. I also suggest that we consequently need to carefully reassess the received wisdom about "reductionism," and "antireductionism," in a variety of ways.

It is also worthwhile emphasizing at the outset that I do not endorse compositional reduction and that elsewhere I have given reasons why it may not ultimately succeed.[6] My primary goal here is simply to establish that a novel, and prima facie plausible, ontological reductionism results once one wholeheartedly follows Kim's strategy of embracing the metaphysics of science, while acknowledging the antireductionist's insights about special sciences My final conclusion is that compositional reductionism poses a challenge that all nonreductive physicalists, as well as "antireductionists" in the philosophy of science, are obliged to meet However, we will also see that the legitimacy, and indispensability, of special sciences does *not* depend upon meeting this challenge

---

[6] See my "Non-Reductive Realization and Non-Reductive Identity," in Sven Walter and Heinz-Dieter Heckmann, eds., *Physicalism and Mental Causation* (Charlottesville, VA: Imprint Academic, 2003), pp. 31–58; "Strong Emergence as a Defense of Non-Reductive Physicalism," *Principia*, VI (2003): 83–114; and "Samuel Alexander's Emergentism," *Synthese*, CLIII (2006): 261–96.

I COMPOSITION IN THE SCIENCES AND AN ENGINE FOR COMPOSITIONAL REDUCTION

Over the last century, scientific investigations have given us detailed, integrated accounts of many natural phenomena through mechanistic explanations that explain one kind of entity using the very different kinds of entity that the sciences take to compose it [7] The neurosciences provide a particularly useful example because they quickly show how comprehensive such compositional relations are For example, consider Figure 1 which illustrates the compositional levels of individuals, and the mechanistic explanations associated with them, involved in the spatial memory of a rat Scientists are increasingly confident that psychological states, like the rat's remembering a certain spatial location, are composed by brain areas, where the latter are themselves composed by collections of cells like neurons, which are in turn composed of cellular structures, like ion channels, which are composed by complex biochemical molecules like protein subunits, although the latter are not shown in the diagram, which are themselves composed by atoms, which are composed by the subatomic entities studied by physics Such compositional hierarchies of individuals, which as we will see are usually accompanied by hierarchies of powers, properties, and mechanisms, and the detailed mechanistic explanations that underpin them, are now found throughout the sciences

To illustrate these notions of composition, I am going to examine a concrete case from the sciences and I need a metaphysical framework to guide my work Since our primary focus will be on properties, I follow recent work in the metaphysics of science and assume a weakened version of Sydney Shoemaker's "causal theory" of properties under which a property is individuated by the causal powers it potentially contributes in this world, under certain conditions, to the individuals in which it is instantiated [8] (I use a related account of the

---

[7] I use 'composition' as a catch-all term for the various relations posited in mechanistic explanations which I examine further below, including parthood and constitution relations between individuals, realization between properties, implementation between processes, and comprising between powers

[8] Shoemaker, "Causality and Properties," in Peter van Inwagen, ed , *Time and Cause* (Boston Reidel, 1980), pp 109–36

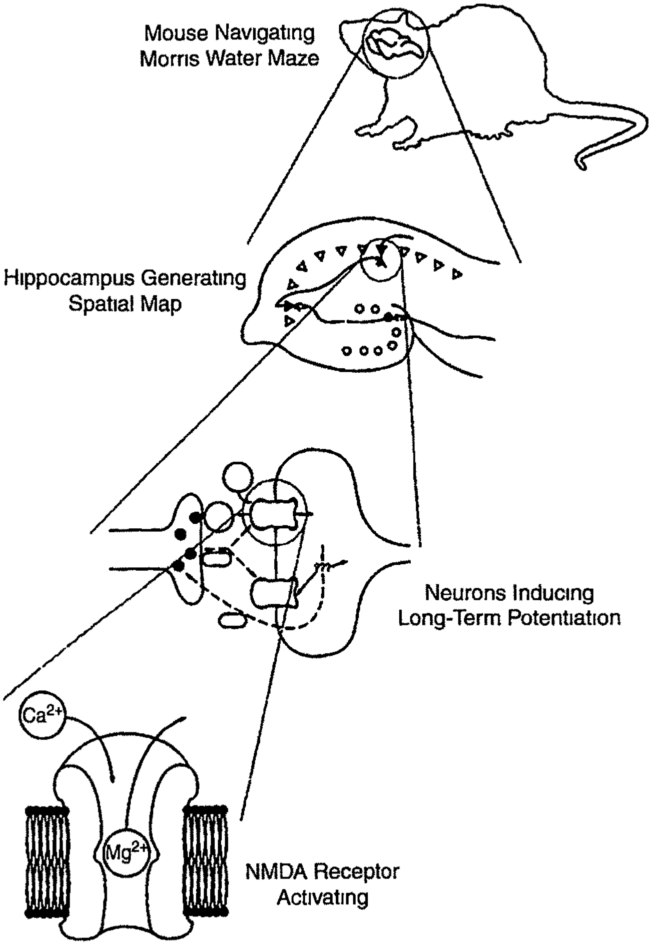


Figure 6 4 Levels in the hierarchical organization of the mechanism of spatial memory

Figure 1 An illustration, using a case from the neurosciences, of the compositional relations typically illuminated by mechanistic explanations in the sciences From Carl F Craver and Lindley Darden, "Discovering Mechanisms in Neurobiology: The Case of Spatial Memory," in Peter K Machamer, Rick Grush, and Peter McLaughlin, eds , *Theory and Method in the Neurosciences* (Pittsburgh University Press, 2001), pp 112–37, on p 118 Reprinted by permission of the University of Pittsburgh Press

individuation of powers which I outline at the relevant point) Following Shoemaker, we should note that the relevant powers may be "forward-looking" powers to cause certain events, or "backward-looking" powers to be caused by specific events For reasons that will become apparent, I am primarily concerned with properties I term *causally efficacious*—properties whose instantiation *actually* determines the contribution of causal powers to an individual[9]

With this machinery in place, consider a prominent example of mechanistic explanation in the neurosciences We know that, under appropriate background conditions, a voltage-sensitive potassium ion channel plays a key role in a neuron due to its property of being a voltage-sensitive gate contributing the backward-looking power of opening in response to a change in the charge of surrounding cells Roderick MacKinnon won the Nobel prize for developing a compelling mechanistic explanation of how such ion channels do this by illuminating the relevant chemical and spatial properties/relations of the complex protein molecules that are "subunits," that is, parts, of these channels[10] Figure 2 offers a textbook illustration of the subunits and Figure 3 provides a simplified illustration of the higher- and lower-level mechanisms whose implementation relations such findings allow us to understand

Basically, when the charge in surrounding cells changes, then the backward-looking powers, to change relative spatial position, of each of the subunits are manifested together and the subunits all swivel to adopt new spatial relations to each other These numerous lower-level mechanisms together implement the higher-level process of the ion channel opening And, in addition, we have a range of other compositional relations, for we have part-whole and constitution relations between the ion channel and protein subunits, so-called "realization" relations between the properties and relations of these individuals, and what I term *comprising* relations between their powers

---

[9] The reader should also note that I use 'entity' as a catch-all referring to powers, properties, individuals, processes, and the like, and that nowhere in the paper will I use the now widely accepted "structural," or "micro-based," properties introduced by D M Armstrong in his *Universals and Scientific Realism* (New York Cambridge, 1978) I have argued in detail elsewhere that once we look at real, concretely described cases of scientific composition we have strong reasons to reject structural properties and analogous entities See my "Why a Naturalistic Diet Is Good for Your (Philosophical) Health" in Brian McLaughlin and Julie Yoo, eds , *New Essays on Realization* (forthcoming)

[10] Below I will focus solely on our mechanistic explanation of the opening of the ion channel, though we also have elegant connected accounts of its other behaviors, again based upon the properties and relations of the subunits For instance, we have such mechanistic accounts of why the ion channel only allows passage to potassium ions, whilst blocking even smaller ions

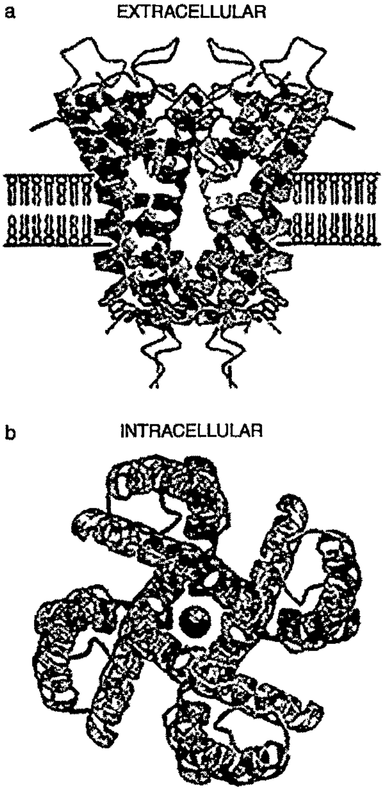


Figure 5–13. Ribbon diagrams to illustrate the three-dimensional structure of the KcsA potassium channel that was solved by Rod MacKinnon and colleagues *a* View from the side Note the "inverted teepee"shape *b* View looking down on the entry to the channel from the extracellular side A potassium ion (*black*) is shown in the central pore formed by the fourfold symmetrical protein (*blue*) See also Plate 5 (Modified from Doyle et al , 1998)

Figure 2 Textbook illustration of the properties and relations of the protein subunits From Irwin Levitan and Leonard Kaczmarek, *The Neuron Cell and Molecular Biology* (New York Oxford 2002), p 111 Modified from D A Doyle et al , "The Structure of the Potassium Channel Molecular Basis of K+ Conduction and Selectivity," *Science*, CCLXXX (1998) 69–77, Figure 3 Reprinted with permission from Oxford University Press and AAAS

To better understand these various compositional notions, let us begin by highlighting a few of their common features First, even a cursory inspection shows that such relations are not a species of causal determination, which is temporally extended, occurs between wholly distinct entities and usually involves the mediation of force and/or

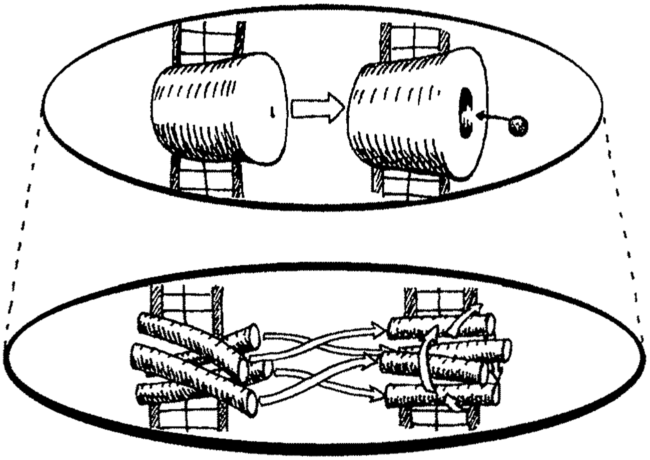


Figure 3 A diagram of the lower-level mechanisms at bottom, involving the protein subunits, that together implement the process of the potassium ion channel opening in response to a change in charge of nearby cells, outlined in the top of the figure

the transfer of energy Rather, the compositional relations posited between powers, properties, individuals, and processes are all synchronous, do not occur between wholly distinct entities and do not involve the mediation of force and/or the transfer of energy I thus use the term *noncausal determination* to refer to the compositional relations involved in mechanistic explanations

Second, we must carefully mark that the entities bearing compositional relations in this case, and the many like it, are usually qualitatively distinct Thus, for instance, the properties and relations of the protein subunits contribute no common powers with the property of the ion channel—the protein subunits have powers such as changing their relative spatial positions, under certain conditions, but not the powers of opening or allowing speedy passage to potassium ions In fact, it makes little sense to ascribe contribution of either the power of opening, or allowing speedy passage to ions, to a property of a protein subunit, for this protein just is not the right type of individual to have such a power or the property that contributes it It is very common for higher-level properties to be ascribed such "novel" powers not had by any of their realizers And, as a quick perusal of any number of examples of mechanistic explanation in the sciences establishes, similar points also hold for the individuating features of the various powers, individuals, and mechanisms involved in such cases—the entities taken to be involved in the relevant compositional relations are usually of *qualitatively different* kinds (Consider again, for example, the differing natures of the entities in Figure 1 that underpin various mechanistic explanations)

It may initially be puzzling how entities of one kind could compose entities of very different kinds and, consequently, it is important to note a final general feature of such compositional relations Given their interconnections, the various "packages" of powers, property instances, individuals, and mechanisms studied by lower-level sciences *together* compose the qualitatively different powers, property instances, individuals, and mechanisms studied by higher-level sciences Compositional relations in the sciences are thus often "many-one" with regard to their relata, with many component entities and one composed entity Not only is this a commonplace feature of scientific composition relations, but it also apparently holds the secret of such relations and the explanations built upon them For although each is qualitatively distinct from the composed entity, the deeper point is that numerous lower-level component entities can *together* noncausally result in the qualitatively different entity they are taken to compose [11]

To further sharpen our understanding of scientific notions of composition, I want to outline briefly precise accounts of the compositional relations posited between properties and powers in mechanistic explanations, starting with the latter since this will serve to highlight the central role of mechanisms and their relations Any successful account of scientific composition must accommodate the general features we have just highlighted and I will shortly argue that such features suffice to drive compositional reduction [12] I thus want to note that compositional reductionism will plausibly work with any adequate view of compositional notions in the sciences and is hence not tied to my own account of realization or any other compositional notion

A power basically is an entity whose possession by an individual allows this individual to enter into certain processes Powers are thus individuated, at least partially, by the particular mechanisms their manifestation grounds But we have just seen that the sciences show that, if manifested, the backward-looking powers of the protein subunits, to change relative spatial positions in response to a change in the charge of surrounding cells, ground mechanisms that together implement the mechanisms grounded, if this power is manifested, by the backward-looking power of the ion channel to open in response to a change in the charge of surrounding cells Thus we can plausibly see that the sciences illuminate how, and why, the powers of the protein subunits comprise the powers of the ion channel, since the mechanisms grounded by the powers of the protein subunits together implement the mechanisms grounded by, and more importantly individuative of, the ion channel's powers Using these observations as a guide, we can frame the following theory schema for comprising

---

[11] In addition to those highlighted above, there are a number of other features of compositional relations For example, such relations usually only hold under certain background conditions, thus the properties and relations of the subunits only realize the properties of the ion channel when, amongst other background conditions, they are encased in the oily membrane of the cell For a fuller survey of all the features of scientific composition see my "Making Sense of Levels in the Sciences" (unpublished)

[12] The reader should mark that elsewhere I have criticized the so-called "subset" or "Flat" view of realization (in "The Dimensions of Realization," *Analysis*, LXII (2001) 316–23), as well as a wide range of other accounts of "making-up" and "composition" in the sciences (in "Why a Naturalistic Diet "), for failing to accommodate the characteristics of scientific composition, particularly the qualitative distinctness of its relata

> (Comprising) Powers C_1–C_n, had by individuals s_1–s_n (or individual s^*), comprise the power C^*, had by individual s^* under background condition $\$$, *if and only if* the mechanisms grounded by the triggering and manifestation of powers C_1–C_n, under triggering conditions $\$t_1$–$\$t_n$ and background condition $\$$, would together implement the mechanisms grounded by the triggering and manifestation of C^*, under triggering conditions $\$t_1$–$\$t_n$ and background condition $\$$, but not vice versa

Similar points hold for realization relations between property instances, since a property instance is an entity that is individuated by the powers it contributes to individuals The sciences again explain why if one has the protein subunits with their properties and relations, then the latter properties will realize the ion channel's property For the subunits' properties and relations together contribute powers that comprise the powers individuative of the property of being a voltage-sensitive gate and hence noncausally result in, and realize, an instance of the property of being a voltage-sensitive gate in the ion channel Building upon our account of comprising, and also remembering that we found the relevant individuals in our case bore parthood/constitution relations to each other, we can offer this theory schema for scientific realization

> (Realization) Property instances F_1–F_n, in individuals s_1–s_n (or individual s^*), realize a property instance G, in individual s^* under background conditions $\$$, *if and only if* the powers contributed by F_1–F_n to s_1–s_n (or s^*), which are constituents/parts of s^*, together comprise the powers individuative of G, in s^* under $\$$, but not vice versa

This account of scientific realization and our schema for comprising refer to each other, and also to parthood and constitution relations between individuals and implementation between mechanisms It should therefore be obvious that to give a comprehensive account of scientific composition we also need to provide theory schemas for parthood and implementation. Elsewhere I have provided such an integrated framework, but my goal here of illustrating scientific composition is amply served by our accounts of realization and comprising.[13]

Given our observations about the general features of scientific composition, as well as our more precise schemas, we can finally turn to debates over "reduction." With regard to earlier discussions, we can immediately understand why so-called "multiple realization" exists, as well as "multiple composition" generally, since distinct combinations of lower-level properties, that are very different from each other, may nonetheless each still *together* noncausally result in instances of the same special science property. Earlier writers, like Putnam and Fodor, argued that such multiple composition means that we rarely get exact coextension between any particular combination of components and a composed entity, thus violating Leibniz's Law and leaving us without identities, where similar points hold with regard to the existence of bridge laws.

In fact, our more detailed account of scientific composition suggests that the difficulties surrounding identities between the entities of higher- and lower-level sciences are even more pronounced. As we just noted, compositional relations are asymmetric, many-one, and relate entities of qualitatively different kinds, among other features. In contrast, identity is always symmetric, one-one, and relates entities of exactly the same kind. As a result, given the differing natures of these relations, the entities related by compositional relations in the sciences are rarely involved in identities and it is plausible that one will not even get so-called "restricted" identities when one does have coextensivity through a "unique" realization base.[14]

---

[13] See my "Making Sense of Levels in the Sciences." Some readers may be concerned that the interdefined nature of the schemas raises problems of circularity. In response, it should be noted that I am offering an account in the metaphysics of science that simply seeks to articulate the nature of compositional concepts in the sciences, rather than trying to replace these scientific notions with other concepts. Furthermore, our basic concepts in some area are often inextricably intertwined and my account merely reflects this feature in the scientific case we have considered. For we have seen that concrete scientific cases posit "packages" of compositional relations between powers, properties, individuals, and mechanisms.

[14] See my "Why a Naturalistic Diet . . . ." Two basic problems recur in securing identities even when we have a unique realization base. First, the differing modal, and other, features of higher- and lower-level scientific entities preclude even "restricted" identities in such cases. And, second, as we noted above, the core characteristics of scientific composition relations contrast sharply with the features of identity relations, whether "restricted" or not, to ensure that we do not usually get identities even when we have unique realization. Such points are bolstered by the rather different critique of such identities offered by Mark Johnston in "Manifest Kinds," this JOURNAL, XCIV, 11 (November 1998): 564–83.

Such difficulties in securing identity statements explain, at least in part, why reductionists have come to regard the Nagelian approach as a liability. And our treatment of scientific composition also allows us to appreciate why Kim's alternative strategy looks far more promising. Furthermore, our work highlights the basis of the ontological reductionism endorsed by the many working scientists who routinely contend that mechanistic explanations show that special science entities are "nothing but," or "nothing more than," the entities taken to compose them.

The key point to note is that with the compositional relations posited in mechanistic explanations the composing entities *together* noncausally result in the composed entity, but not vice versa. An ontological reductionist can therefore argue that we should not be committed to *both* component *and* composed entities. For we can consequently account for the truth of the special science statements commonly understood to refer to composed entities by simply taking the components to be their sole truthmakers. Thus, given our mechanistic explanations, one can, for instance, account for the powers of the ion channel's properties using the powers contributed by the properties and relations of the protein subunits, but not vice versa.

Once one embraces the metaphysics of science, such parsimony arguments allow one to harness the notions of composition deployed in the special sciences in the service of a new ontological reductionism. As I note below, such arguments can be formulated using different premises and can easily be provided for special science powers, individuals, and mechanisms. But, to illuminate such reasoning, I will stick with the recent focus on properties and outline an argument based around realization relations.

Given the nature of realization in the sciences, our reductionist may argue that it is ontologically profligate to take any realized property instance to determine the contribution of causal powers, and hence to be causally efficacious, *in addition* to its realizer property instances. For the reductionist claims that, given the nature of the realization relation, we can account for *all* the causal powers of individuals simply using the contributions of powers by the realizer property instances of these individuals, or their constituents, rather than also as contributions from realized property instances. But we cannot account for all causal powers of individuals simply as contributions by realized property instances. If we assume that the causal powers of individuals are not overdetermined, then, appealing to Occam's Razor, the reductionist argues that we should accept the existence of no more causally efficacious property instances than we need to account for the causal powers of individuals. The proponent of this simple argument thus concludes that we should *only* accept that realizer property instances contribute powers to individuals and hence *only* accept that realizer property instances are causally efficacious But it is also plausibly true that the only property instances that exist are those that make a difference to the causal powers contributed to individuals We may thus further conclude that there are only realizer property instances Let us call this the *Argument from Composition*

From this point on I will use the terms *ontological* or *compositional* reductionism to refer only to the brand of reduction that utilizes the Argument from Composition or analogous parsimony arguments focused on powers, individuals or processes [15] In section II, I examine the wider features of compositional reductionism, but in the remainder of this section I focus on the nature and scope of the Argument from Composition, and similar reasoning We should start by emphasizing that, in the first instance, the Argument from Composition is not focused on *causation*, or causal exclusion principles, or concerns about causal overdetermination, but instead argues to a conclusion about the putative implications of the *compositional* relations between powers and properties posited by mechanistic explanations Second, it should also be clear that, rather than being focused upon any semantic conclusion, such as the adequacy of explanations, projectibility of predicates, or the like, the Argument from Composition is thoroughly ontological in nature, reducing our initial commitments to realizer and realized property instances down to an ontology encompassing only realizers Third, if anything like the general picture of scientific composition broached earlier is correct, then compositional reductionism will extend to reduce our commitments to composed special science powers, individuals, and mechanisms, as well as special science properties For the scientific notions of comprising, constitution or parthood, and implementation are like realization in the relevant component entities being such that together they noncausally result in the composed entity, but not vice versa

How widely one takes this form of reduction to apply in the special sciences, and their component areas, will depend upon one's view of the extent of mechanistic explanations For example, at one extreme, if one believes that, so far, we have only shown in limited patches of the special sciences how the entities posited in these areas are mechanistically explained by lower-level entities taken to compose them, then one will take compositional reductionism only to apply "locally" to this limited group of special science entities At the other extreme, a fully "global" version of reductionism results if one contends that we are justified in believing *all* the entities posited in special sciences will plausibly be mechanistically explicable using entities taken to compose them For instance, proponents of physicalism claim that scientific evidence justifies us in taking the natural world to be a fully comprehensive compositional hierarchy bottoming out, so far as we know, with the entities of physics Since reasoning like the Argument from Composition entails that we should only accept the existence of component entities, the truth of physicalism allows compositional reductionism, if successful, to show that the entities of microphysics are the *only* entities We can thus see that although the applicability of compositional reduction does not depend upon the truth of physicalism, a completely "global" ontological reductionism looms if one accepts this thesis

---

[15] Though I focus here upon a version of compositional reduction based upon the Argument from Composition there are a number of differing formulations of the argument that drives this view Most importantly, one can base compositional reductionism upon what I elsewhere term the "Argument from Composition and Completeness" which takes as its premises *both* the nature of scientific composition *and* the truth of the "Completeness of Physics," that is, the claim that all microphysical events are determined, in so far as they are determined entirely by prior microphysical events See my "Nonreductive Identity and Nonreductive Realization," "Strong Emergence as a Defense ," and "Samuel Alexander's " (The reader should also note that in these papers I argue the Argument from Composition is invalid and that the Argument from Composition and Completeness is valid, but not known to be sound)

At this point, it is useful to contrast compositional reduction with Kim's own formulations of the new reductionism We have already noted one divergence, since compositional reduction explicitly concerns the implications of scientific composition, while Kim has often focused upon causation or concepts of supervenience [16] More importantly, Kim's centerpiece account of "reduction by functionalization," his most detailed treatment of the new reductionism, continues to follow the Nagelian in using identities to drive its ontological reductions—although only identities between property instances [17] However, as we have just noted, the widespread phenomenon of multiple realization in the sciences makes it plausible that we rarely get such identities And a better understanding of scientific composition only reinforces such points by illuminating still deeper difficulties Though the unavailability of such identities obviously hobbles Kim's particular formulation of the new reductionism, such points leave compositional reduction completely untouched given its reliance solely upon realization and other compositional concepts[18]

---

[16] See, for example, the papers in Kim's *Supervenience and Mind* (New York Cambridge, 1993), but even Kim's later *Mind in a Physical World* continues the focus on supervenience

[17] "Reduction by functionalization" is examined in *Mind in a Physical World*, but the most sophisticated treatment is provided in Kim's "Making Sense of Emergence," where the reliance on identities between property instances is also clear (p 15 and elsewhere)

To summarize, we have found that Kim is correct that a more careful examination of composition in the sciences provides a foundation for a philosophically novel form of reductionism, though one familiar to working scientists However, despite its basis in some of their own recent insights about special sciences, many antireductionist philosophers will likely remain unconvinced of the significance of compositional reductionism They will argue that features of scientific practice, such as the widespread scientific acceptance of multiple realization, or the apparent indispensability of special science predicates noted by Fodor in our opening passage, clearly trump any merely metaphysical argument Such responses deserve respect, since they rightly press the reductionist to engage the rich body of work that now exists concerning the nature of scientific practice and its intellectual products In the next section, I therefore examine what compositional reductionism should say about a range of important issues in the philosophy of science

II COMPOSITIONAL REDUCTIONISM, THE SPECIAL SCIENCES, AND THE CONTOURS OF A NEW DEBATE

Compositional reductionism stands recent debates on their heads in various ways and in this section I argue, perhaps unsurprisingly, that it has a character very different from earlier strains of reduction In contrast to Nagelian reductionism, or even Kim's reduction-by-functionalization, I show that compositional reductionism plausibly ought to accept at least the following theses

(a) True identity statements involving the predicates commonly understood to concern special science properties or their instances, and the properties studied by lower-level sciences or their instances, are rare in the sciences
(b) Working scientists illuminate, and endorse, many true sentences that they commonly understand to be about the "realization" and "multiple realization" of special science properties
(c) The predicates of special sciences are projectible

---

[18] It is important to note that, if they exist at all, our commitments to the "structural," or "micro-based," properties feted both by Kim, in *Mind in a Physical World*, and Derk Pereboom, in his "Robust Nonreductive Materialism," this JOURNAL, XCVIII, 10 (October 2002) 499–531, are also consequently reduced by compositional reduction For if such structural properties exist at all, then the sciences make it clear that their powers plausibly noncausally result from the powers of the fundamental microphysical properties, such as the spin, charm, and charge of quarks Consequently, such structural properties are plausibly realized properties and hence an Argument from Composition also applies to them

(d) The special sciences and their predicates provide true explanations
(e) The best explanations in many contexts, given our purposes, are often those using predicates of special sciences
(f) The special sciences and their predicates provide true, law-like generalizations
(g) Many true explanations and/or true law-like generalizations can, in principle, only be properly expressed using the predicates of the special sciences

And

(h) An important part of successful scientific practices of "discovery" is that working scientists often use our common understanding of special science predicates and the sentences based upon them, thus taking these predicates to concern composed entities or relations of composition

This list should be a little shocking, for it includes many claims that have long been disputed by reductionists [19]

To highlight this distinctive character of compositional reductionism, as well as to assess its plausibility, I am going to use as a dialectical foil the two main kinds of argument recently offered against "reductionism" in the philosophy of science, one based around multiple realization and the other upon predicate indispensability I briefly examine these key criticisms in turn and show that in each case the compositional reductionist can embrace the central tenets of such arguments My conclusion is that the extant critiques of "reductionism" fail against compositional reduction Along the way, I also explain why such an ontological reductionism should accept (a)–(h) and I conclude by noting a number of respects in which ongoing discussions of "reduction" need to be reassessed once we appreciate the deeper nature of compositional reductionism

In resisting Nagelian reduction, as we saw in section I, the common antireductionist refrain is that the heterogeneity of lower-level realizers (or "kinds"), that is, the prevalence of multiple realization, precludes “reductionism.” One way of framing the driving idea of these critiques, that will be useful again shortly, is that our world contains many “macro-commonalities,” in features shared by aggregations of microphysical individuals, that are “pure” macro-commonalities—that is, features shared by such aggregations that are microphysically heterogeneous. A pure macro-commonality is thus a commonality found at the macro-level without a corresponding commonality at the micro-level. There are obviously “impure” macro-commonalities, that is, features shared by microphysical aggregations in virtue of their shared microphysical commonalities, but antireductionists have argued convincingly that the Nagelian's mistake is to assume that all macro-commonalities are of this impure type. As we have seen, antireductionists have shown that the prevalence of multiple realization, and accompanying pure macro-commonalities, mean that higher- and lower-level scientific entities, and the predicates we use to refer to them, simply do not neatly line up and thus we can rarely formulate the true identity statements or bridge laws needed for Nagelian reduction.

---

[19] We could also likely add further methodological claims to the list, such as

(i) The sciences which study realized and realizer properties, as commonly understood, intertheoretically constrain each other under appropriate conditions (such as sufficient confirmation of the relevant theories, and so on) and may co-evolve, through a process of mutual fit and adjustment, in a variety of beneficial ways

However, the acceptability of the combination of theses like (i) with (b), and/or (g), introduces a range of further complications I therefore leave theses like (i) to one side here, but for a detailed defense of a position combining (i), (b), and (g) see my paper, co-authored with Kenneth Aizawa, "Multiple Realization and Methodology in the Neurological and Psychological Sciences" (unpublished)

Our work in section I prefigures obvious responses for the compositional reductionist to make to such arguments based around multiple realization. The phenomenon of multiple realization is based upon heterogeneous groups of lower-level realizers each together noncausally resulting in the same powers, in those individuative of some special science property. But, given this point, the compositional reductionist will again argue that we can account for the truth of the statements commonly understood to be about realized properties, whether multiply realized or not, simply by taking the lower-level realizers as their sole truthmakers. The reductionist will also endorse, and account for, the projectibility of special science predicates in a similar fashion. We can thus see that the ontological reductionist ought to endorse theses like (a)–(c), embracing statements commonly understood as concerning multiply realized properties and using their insights to drive her brand of ontological reduction.

Such responses to objections based upon multiple realization still leave important questions unanswered about compositional reductionism and its stance towards special sciences. To engage these issues, let us therefore turn to the second kind of critique of “reductionism” based upon the putative indispensability of special science predicates. To highlight the nature of such arguments, and the compositional reductionist's response to them, I will use work by Kitcher since it fits well with our earlier discussion. However, I contend that the general approach I outline also works with other versions of this argument given by writers like Putnam, Fodor, and Wimsatt.

Kitcher focuses on the case of molecular biology and various higher biological sciences, such as classical genetics, cytology, and embryology He convincingly argues that properties and individuals putatively referred to by predicates in the higher biological sciences are taken to be multiply realized in the case of properties, and what we might term *multiply constituted* in the case of individuals, by heterogeneous molecular properties and individuals Furthermore, Kitcher argues at length that we very often fail in attempts to replace key explanations using the predicates of higher-level sciences, whether of classical genetics, cytology, embryology, or others, with explanations using the predicates of molecular biology To illustrate the thrust of these arguments for my purposes, it suffices to consider the example that Kitcher uses to summarize his wider conclusions Kitcher tells us

> The distribution of genes to gametes is to be explained, not by rehearsing the gory details of the reshuffling of the molecules, but through the observation that chromosomes are aligned in pairs just prior to the meiotic division, and that one chromosome from each matched pair is transmitted to each gamete We may formulate this point in the biologists preferred idiom by saying that the assortment of alleles is to be understood at the cytological level What is meant by this description is that there is a pattern of reasoning which involves predicates that characterize cells and their large scale structures That pattern of reasoning is to be objectively preferred to the molecular pattern which would be instantiated by the derivation that charts that complicated rearrangements of individual molecules because it can be applied across a range of cases which would look heterogeneous from a molecular perspective Intuitively, the cytological pattern makes connections that are lost at the molecular level, and is thus to be preferred (*op cit*, pp 22–23)

Here Kitcher presses the indispensability, in framing certain true explanations, of special science predicates referring to what I earlier dubbed pure macro-commonalities, that is, features that are shared between aggregations of microphysical individuals that are molecularly and hence microphysically heterogeneous in their properties and relations For example, if we solely used the predicates of molecular biology (or microphysics), then Kitcher convincingly argues that we would fail to refer to the macro-commonality shared between molecularly and microphysically heterogeneous aggregations that is central in truly explaining the shared behavior of these aggregations as meiotic processes As a result, special science predicates referring to the relevant pure macro-commonality are, in principle, indispensable in formulating this and many other true explanations

Kitcher summarizes the underlying basis of his argument by referring to what he terms the "organization of nature" which he suggests entails that special science predicates are indispensable in the ways he outlines (*op cit*, p 22) This point allows us to see how the compositional reductionist should respond to arguments like Kitcher's If we focus on the entities of physics as the relevant "components," the compositional reductionist will obviously accept that the entities of physics aggregate, often into vast aggregations involving huge numbers of microphysical individuals with all kinds of microphysical properties and bearing a range of microphysical relations to each other Furthermore, as we have seen, this ontological reductionist will also take such aggregations of microphysical individuals to share pure macro-commonalities, since the compositional reductionist accepts that many true special science statements concern pure macro-commonalities, including those statements usually taken to refer to multiply realized properties Crucially, however, the compositional reductionist will deny that we should reify such aggregations of microphysical individuals or their features, in pure macro-commonalities, into further individuals or properties Instead, drawing on our earlier insights about the general nature of scientific composition, the compositional reductionist will argue that pure macro-commonalities are nothing more than microphysical properties and relations, of aggregations of microphysical individuals, acting together—where, most importantly for our purposes, the reductionist will accept that heterogeneous combinations of microphysical properties and relations may still *act together in the same ways*

Given this stance, the ontological reductionist can adapt and endorse arguments about the indispensability of special science predicates offered by antireductionists such as Kitcher The ontological reductionist will agree that lower-level predicates fail to capture pure macro-commonalities As a result, the compositional reductionist concludes, like Kitcher and other recent antireductionists, that special science predicates, referring to such pure macro-commonalities, are *in principle* indispensable in framing many explanations and other true statements Of course, the key divergence of the compositional reductionist from antireductionists is over the nature of the pure macro-commonalities they both take special science predicates to be necessary in order to describe The compositional reductionist claims that special science predicates are indispensable, in principle, only for referring to the ways of acting together shared by microphysically heterogeneous aggregations—where, using their battery of parsimony arguments, the compositional reductionist argues this commits us only to microphysical properties/relations and individuals, albeit in a variety of heterogeneous aggregations

The latter points show how a proponent of compositional reductionism who defends a thoroughgoing form of ontological reduction may endorse claim (g), and this line of reasoning illustrates reasons why such an ontological reductionist will also endorse claims like (d)–(f).[20] Our work also allows us to diagnose some mistaken assumptions that reductionists have often made about the relation of the predicates of different sciences. One driving idea of many reductionists is that all true scientific statements, whether framed in the vocabulary of physics (or some other lower-level science), or using the different predicates of the special sciences, all have the same referents—the entities of physics (or the relevant component entities). Consequently, perhaps swayed by the Nagelian approach among other influences, reductionists have long assumed the latter point entails that if "reductionism" is true, then there must be true identity statements showing that the predicates of physics (or some other lower-level science) and those of the special sciences have *exactly the same* referents. But our brief account of how the compositional reductionist should respond to arguments based upon predicate indispensability shows that even if our world is structured as an ontological reductionist asserts, then our empirical evidence makes it plausible that such true identity statements will still usually be unavailable. For in such a situation the predicates of physics (or other lower-level science), and those of the special sciences, apparently refer to *different* features of the only entities that there are—in the entities of physics (or the relevant component entities)—so we thus still usually fail to get true identity statements.

These points bring us to other issues surrounding how special science predicates, and the sentences based upon them, are usually understood and the manner in which the compositional reductionists says we should understand them. One can frame the stance of our reductionist using competing frameworks in the philosophy of language and I shall therefore seek to express the core of the position without taking sides in such debates. To this end, I will use the neutral term *ultimate understanding* to refer to the compositional reductionist's favored view of special science sentences under which their predicates are about microphysical properties and relations acting together. For the reasons outlined above, when ultimately understood the compositional reductionist argues that special science sentences often express truths about nature. In contrast, however, working scientists usually have what I am neutrally terming the *common understanding* of special science predicates. That is, in their research scientists routinely, and quite sincerely, take the predicates of special science sentences like 'Ion channels are voltage-sensitive gates' to be about realized special scientific properties such as being a voltage-sensitive gate.

---

[20] Positions combining ontological reductionism and some form of semantic antireductionism have recently been defended in the previously cited papers on new reductionism by Kim, John Heil's *From an Ontological Point of View* (New York: Oxford, 2003), Barry Loewer's "Why Is There Anything Except Physics?" (unpublished), Trenton Merricks's *Objects and Persons* (New York: Oxford, 2001), and van Inwagen's *Material Beings* (Ithaca: Cornell, 1990), amongst others. The machinery of these writers potentially provides a resource for the compositional reductionist, so long as she is careful about the rather different commitments of these writers on a range of issues.

Having roughly distinguished the ultimate and common understandings of special science sentences, we can now see that the compositional reductionist should defend the use of the common understanding of special science predicates in ongoing scientific research given the latter's awesome utility in scientific practices of "discovery." For the reductionist should agree that such practices have illuminated, and continue to illuminate, all manner of truths about the natural world. The compositional reductionist will thus defend thesis (h), though this reductionist obviously contends that when commonly understood special science sentences actually express falsehoods (or fail to refer).[21] Such a stance may appear odd. How could it be a successful methodology for scientists to understand their sentences in a manner which strictly renders them false? However, the compositional reductionist has ready answers about the utility of such a practice under their view.

First, there is mounting evidence from cognitive science that our innate concepts are very much like those involved in what we are terming the common understanding of special science predicates.[22] We thus have a ready account of the special facility of working scientists with special science predicates used with the common understanding. Second, we have already illuminated why the compositional reductionist takes such practices to reveal truths about the natural world. Special science sentences are discovered by scientists using a common understanding of them and under the ultimate understanding of such sentences, articulated by the compositional reductionist, these sentences are very often true—and even express truths that cannot be captured using other predicates. The compositional reductionist may consequently argue that there is no mystery on their view why a common understanding of special science predicates continues to be a successful part of ongoing scientific practices of discovery.

---

[21] The reader should note that I have now committed compositional reductionism to a position about the common understanding of special science sentences at odds with the important alternative approach of van Inwagen's *Material Beings*. Van Inwagen is an ontological reductionist about most individuals, but he argues that our ordinary statements about such individuals are actually true because he suggests a true paraphrase for such statements concerns component individuals (and presumably component properties/relations). Using our earlier findings, a compositional reductionist might potentially avail herself of a similar approach to special science sentences, thus diverging from the strategy outlined in the text.

[22] See, for example, Frank Keil, *Concepts, Kinds, and Cognitive Development* (Cambridge: MIT, 1992).

The differences we have now illuminated between compositional reductionism and previous forms of reduction have important implications for ongoing debates which I briefly examine in the remainder of the paper. Our discussion brings to center stage the fact that there have been two strands in recent debates over "reduction." There are "semantic" claims about entities such as theories, law-like generalizations, explanations, predicates, and so on, like whether or not special science predicates are, in principle, indispensable for expressing certain truths. But there is also another *ontological* strand, too often implicit, which concerns claims about which entities in the world ought to be taken to be the truthmakers for scientific statements. Unfortunately, the dominant Nagelian approach, and its descendents, are centered upon showing how semantic entities of lower-level sciences can do the explanatory work of the semantic entities of higher-level sciences. Under such an approach there is thus a natural tendency to think that if "reduction" is successful, then special science theories, explanations, and predicates must be, at least in principle, dispensable. And this has apparently led to the collapse of the two strands in discussions of reduction, since many philosophers have assumed that they have a dichotomy of options. *Either* be a "reductionist" who endorses some brand of semantic reduction for the special sciences as well as a reduction in our ontology, often to the entities of microphysics alone, *or* be an "antireductionist" who accepts some brand of semantic antireductionism for special sciences and an unreduced ontology encompassing both the entities of physics and "levels" of composed special science entities.

Our earlier work shows the latter to be a *false dichotomy*. The engine of compositional reduction focuses upon the ontological relations between entities posited in the sciences and is hence unlike Nagelian reductionism because it does not reach its ontologically reductive conclusion through any semantic subconclusion about the replaceability, and hence likely dispensability, of special sciences and their vocabularies. It is thus left open to this ontological reductionist, as we have seen, to accept that the predicates of the special sciences not only allow us to frame true explanations and law-like generalizations, but are actually in principle indispensable in capturing many truths about nature.

Understanding the nature of compositional reductionism thus makes it plausible that the central "antireductionist" arguments of the past three decades, based upon multiple realization, predicate indispensability, and related phenomena, only suffice to establish *semantic* antireductionism, for we have seen that an ontological reductionist may embrace such reasoning A pressing question for philosophers who label themselves either "antireductionists," or "reductionists," is consequently what theses such writers should actually endorse and why For example, many "antireductionist" philosophers of science who were implicitly concerned with defending versions of semantic antireductionism, and self-proclaimed "reductionists" who have primarily been committed to establishing the truth of ontological reductionism, need to reassess their stances on, respectively, ontological antireductionism and semantic reductionism

At this point, a common reaction of "antireductionist" philosophers of science is to question whether compositional reductionism poses any challenge to their positions [23] Certainly, they allow, there may be problems in philosophy more broadly about whether nonreductive physicalism, or similar metaphysical views, can be defended against the challenge of ontological reductionism But, such philosophers of science protest, all they ever *really* wanted was to establish the scientifically important conclusion that special sciences are legitimate and/or, in principle, indispensable Since compositional reductionism cedes the latter point, ask these philosophers, why does it pose any problem for "antireductionism" in the philosophy of science or, still more importantly, in the sciences themselves? Such questions are, I suggest, among the most important that now confront us in our new set of debates I will not try to provide a full answer to them here Instead, let me briefly consider a recent dispute in the sciences to show that things are far from being as clear-cut as these "antireductionist" philosophers of science might hope

To this end, consider the recent popular books by two Nobel Prize-winning physicists that reflect wider, ongoing scientific debates in doing battle over the implications of "reduction" and "emergence " In *Dreams of a Final Theory*,[24] the particle physicist Steven Weinberg uses implicit metaphysical arguments to defend ontological reductionism, attack the significance of "emergence," and hence make a case for the special importance, and funding priority, of particle physics. On the other side, Robert Laughlin's book, *A Different Universe: Remaking Physics from the Bottom Down*,[25] emphasizes experimental evidence about "emergence" and the need to fund a range of special sciences pioneering novel approaches to such "emergent" aspects of nature.

---

[23] Scientific realists will confront the compositional reductionist challenge in all the cases where we have sufficient confirmation for the relevant scientific accounts However, many scientific antirealists will also plausibly confront such a challenge For example, a scientific antirealism focused upon "unobservables" will face compositional reductionism in all the scientific cases of "observable" composed entities which have been mechanistically explained in terms of component entities Depending upon one's notion of "observation," such examples still appear to be very numerous indeed

[24] New York Pantheon, 1994

A striking feature of these scientific battles is that, although impressive, the scientific evidence about "emergence" offered by Laughlin and other prominent researchers, such as Philip Anderson[26] or Stuart Kauffman,[27] has made remarkably little headway against the "reductionism" defended by Weinberg and many others, which continues to dominate large areas of the sciences. One possible reason for this state of affairs should now be obvious, for we have seen that a key part of the foundation of one species of reductionism, which I contend is very common among working scientists, is *metaphysical* in nature and largely impervious to further empirical data—for such scientific findings are easily used, in the ways we have outlined, to drive positions like compositional reductionism. Consequently, to defend scientific research and associated methodologies in the special sciences, whether focused upon "emergence" or still more widely, issues which are often close to the hearts of antireductionist philosophers of science, it is imperative to engage compositional reductionism and its arguments.

Against this background, and in combination with our earlier findings, my final conclusion is consequently that Kim's wider contentions hold true and that the focus on Nagelian reduction has indeed led to many unfortunate debates over special sciences that are "beside the point for issues of real philosophical significance." For, to take the most glaring example, we have seen that the strongest form of ontological reductionism actually supports the legitimacy, and in principle indispensability, of special sciences. Nonetheless, just as Kim predicted, compositional reductionism still poses a prima facie plausible challenge that all the nonreductive physicalists in "naturalistic" philosophy more broadly, as well as many "antireductionists" in the philosophy of science, are obliged to address.

CARL GILLETT

Illinois Wesleyan University

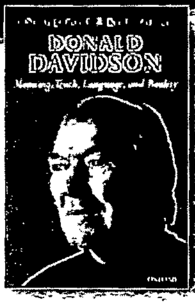
---

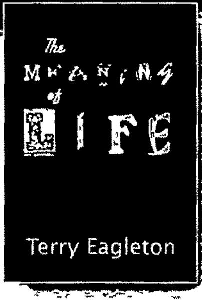
[25] New York: Basic Books, 2005.

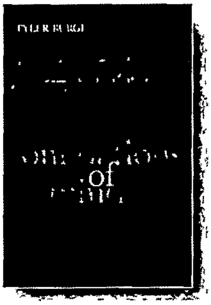
[26] See, for example, Anderson, "More Is Different," *Science*, CLXXVII (1972): 393–96.

[27] Kauffman, *At Home in the Universe* (New York: Oxford, 1995).

New from OXFORD







Donald Davidson
Meaning, Truth, Language, and Reality
ERNEST LEPORE and KIRK LUDWIG
"Lepore and Ludwig have provided a magnificently impressive critical detailed account of Donald Davidsonís contributions to the philosophy of logic, language, and mind "—Gilbert Harman, Princeton University
2007 500 pp paper $39 95 cloth $74 00

Developmental Theism
From Pure Will to Unbounded Love
PETER FORREST
Peter Forrest gives an account of God as a being who has changed from an impersonal God to the God of the Jews, Christians, and Muslims This is not a history of changing ideas about God, but an explanation and defense of the idea of actual "divine development" Forrest uses his theme to explore the intractable concepts of the Problem of Evil, the Trinity, and the Incarnation
2007 216 pp $55 00

The Meaning of Life
TERRY EAGLETON
We have all wondered about the meaning of life But is there an answer? And do we even really know what we're asking? Terry Eagleton takes a stimulating and quirky look at this most compelling of questions at the answers explored in philosophy and literature, at the crisis of meaning in modern times, and suggests his own solution to how we might rediscover meaning in our lives
2007 192 pp, 10 b/w illus $19 95

Foundations of Mind
TYLER BURGE
Foundations of Mind collects the essays which established Tyler Burge as a leading philosopher of mind This second volume of his papers offers nineteen pieces published between 1975 and 2003, including the influential series that develops anti-individualism Burge contributes three essay-length postscripts, a substantial new paper on consciousness, and an introduction which surveys his work in this area The foundations that Burge reflects on are conditions in the individual or the wider world that determine the natures of mental kinds The conditions include causal, social, psychological conditions, and conditions of phenomenal consciousness
2007 456 pp cloth $99 00 paper $35 00

**CHOICE* Outstanding Academic Book 2006*

Now in paperback!
The Stoic Life
Emotions, Duties, and Fate
TAD BRENNAN
"Brennan's book is an unusual specimen a scholarly book that is hard to put down Highly recommended "
—*CHOICE*
A witty, lively, and accessible demonstration of the continuing importance and relevance of Stoicism
2007 352 pp paper $27 50 cloth $45 00

OXFORD
UNIVERSITY PRESS

Prices are subject to change and apply only in the US
To order, please call 1-800-451-7556 In Canada, call 1-800-387-8020
Visit our web site at www oup com/us

THE JOURNAL OF
PHILOSOPHY

BACK ISSUE SALE

Complete volumes
are now available
at the following discounted prices
in order to reduce our inventory:

Volumes 1 through 98 $10 each
Volumes 99 through 101 $25 each

For more information, contact Pamela Ward,
Business Manager, at pcw1@columbia.edu

Visa, Mastercard, American Express and
PayPal are accepted through our website at
www.journalofphilosophy.org

The Conversation of Humanity

STEPHEN MULHALL

"Mulhall is one of the most productive philosophers of his generation within, and contesting, the Anglo-American tradition of philosophy. I should say plainly that I find the brief account of my work on moral perfectionism (which serves as a point of orientation for the lectures as a whole) to be remarkable." **—Stanley Cavell**

"No one around today writes with such elegance, power, and clarity about these difficult authors; there is always about a Mulhall text an almost hallucinatory lucidity."**—Jay M. Bernstein, The New School, author of *Classical and Romantic German Aesthetics***

Page-Barbour Lectures
$25.00 cloth

The Making of Modern Cynicism

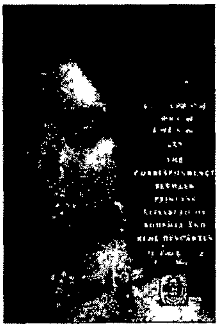
DAVID MAZELLA

"*The Making of Modern Cynicism* is an encyclopedic story of semantic and cultural sea change. Mazella shows how the 'Cynicism' of the ancient world, represented by the shadowy but powerful figure of Diogenes, turns into the watered-down 'cynicism' of the present. Along the way are some surprises. Oscar Wilde is predictably present, but so are Beau Brummell and even Jane Austen. An eye-opening account."**—W. B. Carnochan, Emeritus, Stanford University**

$35.00 cloth

UNIVERSITY OF VIRGINIA PRESS
800-831-3406
www.upress.virginia.edu

European Philosophers



New from Chicago

THE CORRESPONDENCE BETWEEN PRINCESS ELISABETH OF BOHEMIA AND RENÉ DESCARTES

Edited and Translated by Lisa Shapiro

Between the years 1643 and 1649, Princess Elisabeth of Bohemia (1618–80) and René Descartes (1596–1650) exchanged many letters Their correspondence contains the only known extant philosophical writings by Elisabeth, revealing her mastery of metaphysics, analytic geometry, and moral philosophy

THE OTHER VOICE IN EARLY MODERN EUROPE

PAPER $25 00

New from Reaktion Books

LUDWIG WITTGENSTEIN

Edward Kanterian

Ludwig Wittgenstein's personal life, work, and his historical moment intertwined in a fascinating, complex web Noted scholar Edward Kanterian explores these intersections in *Ludwig Wittgenstein* and proposes thought-provoking new interpretations of Wittgenstein's published work

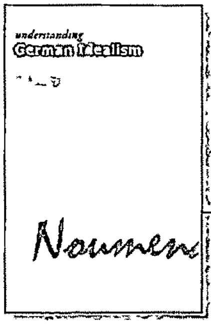
PAPER $16 95

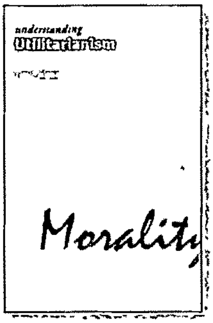
THE UNIVERSITY OF CHICAGO PRESS • www press uchicago edu

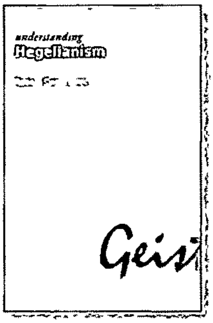
You can now subscribe to the *Journal of Philosophy* using

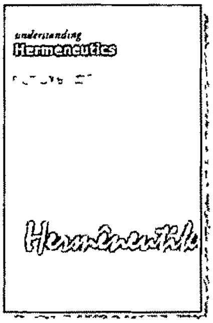
PayPal

To learn more about this easy new option, click on the PayPal icon at **www.jphil.org.**









THE LATEST FROM THE

Understanding Movements in Modern Thought Series

Understanding German Idealism
Will Dudley
978-1-84465-096-5 $22.95 paper
978-1-84465-095-8 $85.00 cloth

An accessible introduction to a complex philosophical movement that revolutionized philosophy and had a tremendous impact on the humanities and social sciences.

Understanding Utilitarianism
Tim Mulgan
978-1-84465-090-3 $22.95 paper
978-1-84465-089-7 $85.00 cloth

"A superb book, wonderfully engaging, philosophically up-to-date and sophisticated." – Brad Hooker, University of Reading

Understanding Hegelianism
Robert Sinnerbrink
978-1-84465-094-1 $22.95 paper
978-1-84465-093-4 $85.00 cloth

A clear and lucid guide to one of the richest and most complex movements in modern philosophy.

Understanding Hermeneutics
Lawrence K. Schmidt
978-1-84465-077-4 $22.95 paper
978-1-84465-076-7 $85.00 cloth

"A very clear, concise, and informative survey of contemporary philosophical hermeneutics and an accessible introduction to the issues for the student and general reader."
– Qingjie Wang, Chinese University of Hong Kong

MQUP | McGILL-QUEEN'S UNIVERSITY PRESS www.mqup.ca

crítica

REVISTA HISPANOAMERICANA DE FILOSOFÍA

Artículos

STEPHEN P TURNER *Practice Relativism*

LIZA SKIDELSKY *La distinción doxástico-subdoxástico*

DAN LÓPEZ DE SA *The Chief Supreme Court Justice A Metaphysical Puzzle?*

Discusiones

JOSÉ E CHAVES RUIZ *Un pseudoproblema para el minimismo*

RICARDO SALLES *Necesidad y lo que depende de nosotros Sobre la interpretación de Marcelo Boeri del compatibilismo estoico*

MARCELO D BOERI *Necesidad, lo que depende de nosotros y posibilidades alternativas en los estoicos Réplica a Ricardo Salles*

Notas bibliográficas

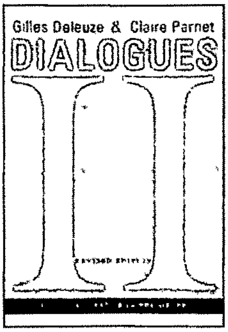
JOSÉ MARÍA TORRALBA, *Acción intencional y razonamiento práctico según G E M Anscombe* y G E M ANSCOMBE, *La filosofía analítica y la espiritualidad del hombre* [Gustavo Ortiz-Millán]

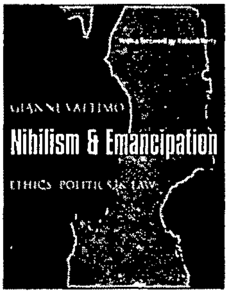
Libros recibidos

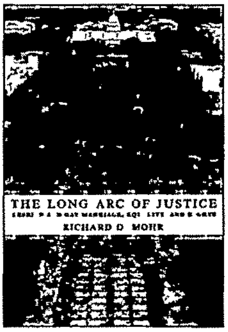
Anexo: 40 años de *Crítica*

Vol 39 / No 115 / abril 2007 / ISSN 0011-1503

Crítica, Revista Hispanoamericana de Filosofía aparece tres veces al año en los meses de abril, agosto y diciembre Es una publicación del Instituto de Investigaciones Filosóficas (UNAM) Todo tipo de correspondencia deberá ser dirigida al Apartado Postal 70-447, Coyoacán, 04510, México, D F Correo electrónico critica@filosoficas unam mx
Versión electrónica http //critica filosoficas unam mx







C O L U M B I A, *Refer to us*

Read book excerpts at www.columbia.edu/cu/cup

Dialogues II

Revised Edition

Gilles Deleuze & Claire Parnet

"The best introduction to Deleuzian philosophy."— Eric Alliez, author of *Signature of the World*

paper - $24.50 European Perspectives

The Philosophy of Qi

The Record of Great Doubts

Kaibara Ekken

"A major contribution to Neo-Confucian thought." — T. U. Weiming, *Harvard University*

cloth $34.50 - Translations from the Asian Classics

Action, Art, History

Engagements with Arthur C. Danto

Edited by Daniel Herwitz & Michael Kelly

"This collection of very lively, eminently readable essays provides a most satisfying intellectual portrait." — David Carrier, *Cleveland Institute of Art*

cloth $45.00 - Columbia Themes in Philosophy

NEW IN PAPER

Animal Rights and Moral Philosophy

Julian H. Franklin

paper $20.00

Nihilism and Emancipation

Ethics, Politics, and Law

Gianni Vattimo

paper $18.00 European Perspectives

The Long Arc of Justice

Lesbian and Gay Marriage, Equality, and Rights

Richard D. Mohr

paper - $16.95

AVAILABLE FROM EDINBURGH UNIVERSITY PRESS

Understanding Ethics

An Introduction to Moral Theory

Torbjorn Tannsjo

paper $23.50

ANNOUNCING:

THE JOURNAL OF PHILOSOPHY

ONLINE

The Journal of Philosophy has added
a print and online subscription option for institutions

SUBSCRIPTION OPTIONS

Libraries and Institutions (print and online)	$180.00
Libraries and Institutions (print only)	$100.00
Individuals	$45.00
Students, retired/unemployed philosophers	$20.00
Postage outside the U.S.	$15.00

Please consult our website at www.journalofphilosophy.org
for more details

20 10 07

THE JOURNAL OF PHILOSOPHY

VOLUME CIV, NUMBER 5
MAY 2007

page
217 *Rational Choice, Risk Aversion, and Evolution* Samir Okasha
236 *The Long Road to Skepticism* Baron Reed

BOOK REVIEWS
263 Gabriel Richardson Lear *Happy Lives and the Highest Good An Essay on Aristotle's Nicomachean Ethics* Iakovos Vasiliou

Published by **The Journal of Philosophy, Inc.**

THE JOURNAL OF PHILOSOPHY

FOUNDED BY FREDERICK J E WOODBRIDGE AND WENDELL T BUSH

Purpose To publish philosophical articles of current interest and encourage the interchange of ideas, especially the exploration of the borderline between philosophy and other disciplines

Editors Bernard Berofsky, Akeel Bilgrami, John Collins, Arthur C Danto, Kent Greenawalt, Patricia Kitcher, Philip Kitcher, Isaac Levi, Wolfgang Mann, Mary Mothersill, Frederick Neuhouser, Christopher Peacocke, Carol Rovane, Achille C Varzi *Consulting Editors* David Albert, James T Higginbotham, Robert May, Charles D Parsons, Wilfried Sieg *Managing Editor* John Smylie

THE JOURNAL OF PHILOSOPHY is owned and published by the Journal of Philosophy, Inc *President,* Arthur C Danto, *Vice President,* Akeel Bilgrami, *Secretary,* Daniel Shapiro, *Treasurer,* Barbara Gimbel, *Other Trustees* Lee Bollinger, Leigh S Cauman, Kent Greenawalt, Michael J Mooney, Lynn Nesbit

All communications to the Editors and Trustees and all manuscripts may be sent to John Smylie, Managing Editor, Mail Code 4972, 1150 Amsterdam Avenue, Columbia University, New York, New York 10027 FAX (212) 932-3721

You may also visit our website at www journalofphilosophy org

THE JOURNAL OF PHILOSOPHY

2007

SUBSCRIPTIONS (12 issues)

Individuals	$45 00
Libraries and Institutions	$100 00
Libraries and Institutions (print and online)	$180 00
Students, retired/unemployed philosophers	$20 00
Postage outside the U S	$15 00

Payments only in U S currency on a U S bank All back volumes and separate issues available back to 1904 Please inquire for advertising rates, ad space is limited, so ad reservations are required

Published monthly as of January 1977, typeset by Dartmouth Journal Services, Waterbury, VT, and printed by The Sheridan Press, Hanover, PA

All communication about subscriptions and advertisements may be sent to Pamela Ward, Business Manager, Mail Code 4972, 1150 Amsterdam Avenue, Columbia University, New York, NY 10027 (212) 866-1742

All materials published in *The Journal of Philosophy* are copyrighted by the *Journal* and are protected by U S copyright law All rights reserved For more information about our copyright policies, please consult our January issue or contact us

The Journal of Philosophy (ISSN 0022-362X) is published monthly by The Journal of Philosophy, Inc , MC4972, Columbia University, 1150 Amsterdam Avenue, New York, NY 10027 Periodicals postage paid at New York, NY, and other mailing offices

POSTMASTER Please send address changes to *The Journal of Philosophy* at MC4972, Columbia University, 1150 Amsterdam Avenue, New York, NY 10027

© Copyright 2007 by the Journal of Philosophy, Inc ISSN 0022-362X

University

THE JOURNAL OF PHILOSOPHY
VOLUME CIV, NO. 5, MAY 2007

RATIONAL CHOICE, RISK AVERSION, AND EVOLUTION*

Suppose you are offered a choice between two options, A and B. Option A gives you \$5 for sure, option B gives you either \$10 or nothing, depending on the flip of a coin that you know to be fair. Which option should you prefer? According to standard expected utility theory, you should be indifferent between the two options, presuming your utility function is linear in money. For A and B have the same expected value, namely five dollars—since $E(B) = 1/2$ $(\$10) + 1/2$ $(\$0) = \$5 = E(A)$, where '$E$' denotes the expected value function. The rational agent, who strives to maximize her expected utility, should therefore be indifferent between the two options.

It is a well-known fact, confirmed by much experimental evidence, that people are usually *not* indifferent between options of this sort—they tend to prefer A to B.[1] This is often expressed by saying that people are *risk averse*—they generally prefer \$$x$ to any gamble with an expected payoff of \$$x$, in apparent violation of the expected utility principle. Economists and decision theorists have long puzzled over how the phenomenon of risk aversion can be reconciled with rational choice theory (of which the expected utility principle is a key part). One attempted reconciliation, due originally to Daniel Bernoulli, argues that utility is *not* a linear function of money but rather is concave, that is, money has diminishing marginal utility. Ten dollars does not bring twice as much utility as five, on this view, so preferring option A

---

* Thanks to Brian Skyrms, Peter Godfrey-Smith, Alan Hajek, and Michael Strevens for comments and discussion, to an anonymous referee for suggestions, and to audiences at the Universities of Bristol, Lancaster, Durham, and the Australian National University, where versions of this paper were presented. Thanks to the Leverhulme Trust for financial support.

[1] For a recent discussion, see Matthew Rabin and Richard H. Thaler, "Risk Aversion," *Journal of Economic Perspectives*, xv, 1 (2001): 219–32.

0022-362X/07/0405/217–35

to *B* does not violate expected utility maximization at all Though it is certainly plausible that money has diminishing marginal utility, many theorists now believe that this cannot explain the observed extent of risk-averse behavior among humans [2] This has led to alternative theories of decision making under risk that reject the expected utility principle altogether, the "prospect theory" of Daniel Kahneman and Amos Tversky is a well-known example [3]

In this paper I want to suggest a new way of thinking about the phenomenon of risk aversion based on evolutionary considerations, and to discuss its philosophical implications My suggestion exploits a line of argument employed by Brian Skyrms[4] in his recent work on the evolution of the social contract I start by briefly describing Skyrms's approach, I then apply his approach to the problem of risk aversion I show that risk-averse behavior is selectively advantageous under many circumstances, so may have an evolutionary explanation Next I discuss the implications of this hypothesis for the traditional debate over how to understand risk aversion, I argue that it supports the view that aversion to pecuniary risk *does* conflict with expected utility maximization Finally, I ask what if anything this shows about the rationality of behaving in a risk-averse way

I SKYRMS ON THE RELATION BETWEEN RATIONAL CHOICE AND EVOLUTION

Skyrms argues that there is a deep analogy between rational choice theory, particularly as it applies to games of strategy, and evolutionary theory In a standard rational choice scenario, an agent is faced with a choice between a set of options, the aim of the theory is to say which choice is optimal As Skyrms notes, it is easy to transpose such a scenario to an evolutionary context Instead of thinking of a conscious agent trying to choose between the options, we can think of natural selection as doing the choosing, favoring the option with the greatest Darwinian fitness Just as the rational person favors the option that maximizes their expected utility, according to traditional rational choice theory, so natural selection favors the option that confers the greatest expected reproductive success on its bearer Expected utility is thus analogous to expected number of offspring, the maximization of the former by rational agents is analogous to the maximization of the latter by natural selection. Just as rational choice theorists argue that much human behavior can be understood as an attempt to maximize expected utility, so evolutionary theorists argue that much animal behavior can be understood as an attempt to maximize reproductive output.

---

[2] See, for example, Rabin, "Risk Aversion and Expected Utility Theory A Calibration Theorem," *Econometrica*, LXVIII (2000) 1281–92

[3] Kahneman and Tversky, "Prospect Theory An Analysis of Decision under Risk," *Econometrica*, XLVII (1979) 263–91

[4] See Skyrms, *The Evolution of the Social Contract* (New York Cambridge, 1995), "Game Theory, Rationality, and Evolution of the Social Contract," *Journal of Consciousness Studies*, VII, 1–2 (2000) 269–84, *The Stag Hunt and the Evolution of Social Structure* (New York Cambridge, 2004)

Skyrms is not the first person to note this analogy, as he acknowledges, but he develops it in an ingenious way. One of his key ideas is that the analogy, although deep, does not hold in every respect. In particular, natural selection can lead to the evolution of strategic behaviors that, in a rational choice setting, would never be adopted by a utility-maximizing agent. Skyrms illustrates this point with a number of examples, the most striking of which is cooperation in the one-shot Prisoner's Dilemma (PD). Consider a standard PD game, first of all in a rational choice rather than an evolutionary setting. Agents have two options, cooperate (*C*) or defect (*D*), with a payoff matrix as shown below; assume that the payoffs are in units of utility.

	C	*D*
C	(6, 6)	(0, 10)
D	(10, 0)	(2, 2)

Payoffs for (*Player* 1, *Player* 2)

Prisoner's Dilemma

What should the rational agent do? As is widely known, if the game is played only once then the best option is to defect. For defecting *strongly dominates* cooperating: whatever one's opponent does, one gets a greater payoff by defecting than cooperating (10 versus 6 if one's opponent cooperates, 2 versus 0 if he defects). Given that the players' choices cannot causally influence each other, defection is thus the rational strategy.

Skyrms expresses this well-known point slightly differently[5]; he says that a Bayes-rational agent, who seeks to maximize expected utility, should never choose a strongly dominated option—hence defection in the one-shot PD follows from the injunction to maximize expected utility. In saying this, Skyrms is assuming the truth of *causal decision theory* (CDT), that is, that in calculating expected utility, we should use the probability that a certain outcome *would* occur if a certain act were chosen (rather than the conditional probability of the outcome, given the act) [6] As we shall see, the CDT assumption plays a crucial role in his argument

---

[5] See Skyrms, *The Evolution of the Social Contract*, chapter 5, and "Game Theory, Rationality and Evolution of the Social Contract," pp. 270–71.

How do we transpose the Prisoner's Dilemma from a rational choice to an evolutionary setting? The trick is to imagine a large population of organisms who interact in pairs, the outcomes of the interactions affect how many offspring they leave Organisms are of two types, cooperators and defectors, as in the standard PD game The payoff matrix is exactly as above, except that the payoffs are now measured in units of offspring, rather than utility So an organism's payoff determines how many offspring it leaves, that is, its reproductive output Following Skyrms, it is assumed that reproduction is asexual and that strategies "breed true", that is, cooperators give rise to offspring who cooperate, and similarly for defectors Given these assumptions, the payoff matrix allows the population's evolutionary dynamics to be studied We can ask questions such as Will the proportion of cooperators in the population increase or decrease, from a given starting point? Can a population of cooperators be invaded by a single defector? Will the population evolve to an equilibrium state where only one strategy is present?

To answer these questions, one further crucial piece of information is needed We need to know the pattern of interaction in the population, that is, the pairing rules One possibility is that pairs are formed at random, that is, every organism has the same probability of being paired with every other Random pairing is the assumption made by most evolutionary game-theoretic models, mainly for reasons of mathematical simplicity [7] Skyrms shows that *if* random pairing is assumed, then defecting has greater fitness than cooperating, that is, defectors leave more offspring than cooperators, on average, so the proportion of defectors in the population increases every generation It is intuitively clear why this is so Random pairing means that the probability of being paired with a *C* is the same for both *C* and *D* types, similarly for the probability of being paired with a D Therefore, the fitness of type C equals

---

[6] These two probabilities—the probability of the counterfactual, and the conditional probability—are often the same, but diverge in Newcomb-like cases (cf Allan Gibbard and William L Harper, "Counterfactuals and Two Kinds of Expected Utility," in C A Hooker, J L Leach, and E F McClennen, eds , *Foundations and Applications of Decision Theory, Volume 1* (Boston Reidel, 1978), pp 125–62) If expected utility were calculated from conditional probabilities, Newcomb-like cases would be counterexamples to Skyrms's claim that expected utility maximizers should never choose strongly dominated options

[7] With random pairing, the frequency distribution of pair-types playing the game on any one occasion is binomial, that is, if p is the proportion of cooperators in the population, then the proportions of the (C, C), (C, D) and (D, D) pairs are p^2, $2p(1-p)$ and $(1-p)^2$ respectively So random pairing means that the frequency distribution of pair-types can be deduced from the frequency distribution of strategies and vice-versa Other pairing rules lack this convenient feature

> 6(probability of being paired with a C) + 0(probability of being paired with a D),

while the fitness of type D equals

> 10(probability of being paired with a C) + 2(probability of being paired with a D)

Clearly, the fitness of type D is greater than that of type C, so the proportion of defectors must increase every generation, whatever the population's initial composition Natural selection eliminates the nice guys

What this shows is that with random pairing, evolution and rational choice coincide In the traditional rational choice setting, defection is the optimal strategy in the one-shot PD, as we have seen Transposed to an evolutionary setting, where payoffs are measured in units of offspring rather than utility, defection is the fittest strategy, given random pairing The analogy between rational choice and evolution therefore works nicely Just as rational agents should defect rather than cooperate in the one-shot PD, given that defecting maximizes expected utility, so natural selection should lead to the evolution of defection in one-shot PD type-situations, given that defecting maximizes expected number of offspring In other words, the optimal strategy in the rational choice sense of 'optimal' is the same as the optimal strategy in the evolutionary sense of 'optimal'

However, there is a complication For as Skyrms shows, this reasoning depends crucially on the assumption of random pairing in the evolutionary case If we drop that assumption, then evolution and rational choice part ways In particular, if pairing is *correlated* then cooperators may have a higher fitness than defectors, so cooperation may be favored by natural selection Correlated pairing means that cooperators are more likely to find themselves paired with other cooperators, and similarly for defectors (Such a correlation may arise for various biological reasons [8]) It is easy to see why correlated pairing can permit the evolution of cooperation Any organism, whether it is a defector or cooperator, gains an advantage from being paired with a cooperator With random pairing, this advantage falls equally on cooperators and defectors, but with correlated pairing it falls preferentially on other cooperators If the correlation is sufficiently strong, this may offset the advantage enjoyed by defectors in mixed (*C*, *D*) interactions, giving cooperators a higher fitness overall.

---

[8] For a discussion of some of these reasons, see Elliott Sober and David Sloan Wilson, *Unto Others The Evolution and Psychology of Unselfish Behavior* (Cambridge Harvard, 1998), Part 1

What exactly does this show? Skyrms's own discussion suggests two main morals. Firstly, the analogy between rational choice and evolution is more complicated than it seems at first sight; the closeness of the analogy depends on the assumptions of the evolutionary model in question. In the one-shot PD with random pairing, the analogy goes through nicely—the strategy that the rational agent favors is the same as the strategy that evolution favors. With correlated pairing this is not so; evolution can favor a strategy—cooperation—that the rational agent should never prefer, for it is strongly dominated. As Skyrms says, with correlated pairing, "rational choice theory completely parts ways with evolutionary theory. Strategies that are ruled out by every theory of rational choice can flourish under favourable conditions of correlation."[9] This parting of ways is particularly striking given that the assumption of random pairing is usually made for mathematical convenience, rather than for empirical reasons.

Note that Skyrms's argument at this point rests essentially on the CDT assumption. Correlated pairing, in the evolutionary PD, means that an organism's strategy becomes probabilistically dependent on its opponent's. An analogous probabilistic dependence can arise in the rational choice PD—for example, if each player believes that his opponent is psychologically similar to himself, and thus likely to play the same strategy that he does.[10] But this does *not* make it rational to cooperate, according to CDT, since the two players' choices are still *causally* independent, which is what matters. However, anyone who rejects CDT, that is, who calculates expected utility using conditional probabilities rather than probabilities of counterfactuals, *would* recommend cooperation in such a case, thus restoring parity with the evolutionary PD. So when Skyrms says that correlated pairing in the one-shot PD leads evolution and rational choice to part ways, his assertion is only true given causal decision theory. This is not necessarily a problem, given the widespread acceptance of CDT, but it is nonetheless important.

The second moral that Skyrms draws concerns the explanation of human behaviors that cannot be squared with rational choice theory. It is well known that humans tend to cooperate, or behave "nicely," much more than considerations of pure rational choice would predict. Indeed, there is substantial experimental evidence that in the one-shot Prisoner's Dilemma (and related games) cooperative strategies are played relatively often, even though such strategies do not maximize expected utility.[11] In the light of Skyrms's work, it is natural to ask whether evolutionary considerations may explain these behavioral anomalies. For humans are an evolved species, and Skryms has shown how under certain conditions, strategies ruled out by rational choice theory can nonetheless evolve. So if we accept that human behavioral tendencies are in part the product of evolution,[12] the fact that humans sometimes behave in ways that are at odds with rational choice, for example, cooperating in a one-shot PD, becomes potentially explicable. Obviously, detailed empirical work would be needed to substantiate the claim that such behaviors have an evolutionary explanation. Skryms does not provide this, but his demonstration of how evolution and rational choice can part ways is an important first step.

---

[9] Skyrms, *The Evolution of the Social Contract*, p. 106.

[10] This is the famous case of the one-shot Prisoner's Dilemma played against a psychological clone; see David Lewis, "Causal Decision Theory," *Australasian Journal of Philosophy*, LIX (1981): 5–30.

In what follows, I use the phenomenon of risk aversion to illustration a further parting of ways between rational choice and evolution, which is in some ways more radical than the ones considered by Skryms. This in turn raises the prospect of an evolutionary explanation for why humans systematically exhibit risk-averse behavior, in apparent violation of expected utility theory.

II. RISK AVERSION IN AN EVOLUTIONARY CONTEXT

Recall the problem we started with: whether a person should choose option *A*, which guarantees a payoff of $5, or option *B*, which pays either $10 or nothing, each with probability 1/2. We noted that utility-maximizing agents should be indifferent between the two options, assuming their utility function is linear in money, for both have an expected payoff of $5. Let us grant the linearity assumption for the moment, equivalently, assume that the payoffs are measured directly in units of utility, rather than money. Expected utility theory thus speaks unambiguously: the rational agent should be indifferent between *A* and *B*.

We need to transpose this problem from a rational choice to an evolutionary setting, just as Skyrms does for the one-shot Prisoner's Dilemma. The transposition is straightforward. We simply let *A* and *B* represent different types of organism in a biological population, as before, assume that reproduction is asexual and that types breed true Type A organisms always produce 5 offspring, type Bs produce either 10 offspring or none, each with probability 1/2 Thus the reproductive output of an A type is fixed, while that of a B type varies stochastically Plausible biological scenarios that could generate such stochastic variation are numerous[13] For example, B types might forage in an area containing lots of high-quality resources but also lots of predators, A types in a resource-poor area with no predators So B types are gambling, A types playing it safe

---

[11] See, for example, E. Fehr and J. Henrich, "Is Strong Reciprocity a Maladaptation?" in Peter Hammerstein, ed., *Genetic and Cultural Evolution of Co-operation* (Cambridge: MIT, 2003), pp. 55–82.

[12] Importantly, accepting this does *not* mean accepting that human behavior is genetically determined. As Skyrms rightly stresses, the assumption that strategies "breed true," that is, are transmitted faithfully from parent to offspring, does not mean that they are under genetic control, one can just as well assume that offspring learn from their parents, or imitate them.

Clearly, the expected number of offspring of an A organism is the same as that of a B organism—five For $E(B)$ = 1/2 (10 offspring) + 1/2 (0 offspring) = 5 offspring = $E(A)$, where 'E' denotes the expectation function (Note that this calculation has exactly the same form as the calculation of expected utility in the analogous rational choice case) This means that on average, As and Bs will leave the same number of offspring One *might* conclude from this that natural selection will not discriminate between them, that is, that neither type will tend to spread at the expense of the other If this were so, then rational choice and evolution would coincide, just as in the Prisoner's Dilemma with random pairing Rational choice theory says that A and B are equally good options, for their expected utility is the same, evolutionary theory *seems* to say that A and B are equally fit, for their expected reproductive output is the same So the preference of the rational agent seems to coincide with the "preference" of the evolutionary process

However, this reasoning rests on two assumptions (i) that the population is very large[14], and (ii) that the reproductive output of each type B organism is independent of every other If either assumption is relaxed, then one cannot infer that selection will not discriminate between types A and B, simply because they have the same expected number of offspring Evolutionary theorists have long known that if there is stochastic variation in offspring number, expected number of offspring may not be a suitable measure of Darwinian fitness—and will not be if the population is small, or if the independence assumption above is relaxed[15] In a small population, or in a population

---

[13] For a good discussion, see J Seger and H J Brockmann, "What Is Bet-Hedging?" *Oxford Surveys in Evolutionary Biology*, iv (1987) 182–211

[14] "Very large" means "large enough so that the average reproductive success of a type deviates negligibly from its expected reproductive success "

[15] See J Gillespie, "Natural Selection for Variances in Offspring Number A New Evolutionary Principle," *American Naturalist*, cxi (1977) 1010–14, and S Frank and M Slatkin, "Evolution in a Variable Environment," *American Naturalist*, cxxxvi (1990) 244–60

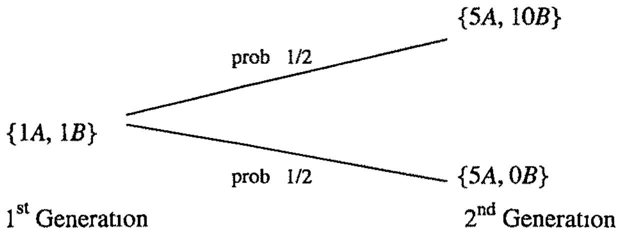


Figure 1 Evolutionary dynamics of a two-membered population across one generation

where independence fails, then if two types have the same expected reproductive output, the type with the lower *variance* in reproductive output will have a fitness advantage, that is, will be more likely to spread (Strictly this is true for any finite population, but the advantage becomes negligible as the population size grows very large) Therefore in many realistic biological scenarios, the "risk averse" A type will enjoy an advantage over the "risk seeking" B type, *despite* their expected reproductive output being identical Put differently, average performance is not all that matters to evolution—variation in performance matters too

An example may help clarify why this is Imagine for simplicity that a population contains just two organisms, one A and one B The relative frequency of the A type in the population is 1/2, obviously, and similarly for the B type Consider the evolutionary dynamics across one generation Given the stochastic variability in B's output, we cannot deduce the population's composition in the next generation, but we can deduce the probability distribution over the possible compositions With probability 1/2, the population will contain $5A$ and $10B$, and with probability 1/2 it will contain $5A$ and $0B$ (see Figure 1) So the expected *number* of A organisms, after one generation, is the same as the expected number of B organisms, namely 5 Crucially, however, the expected *frequency* of the A type, after one generation, is greater than that of the B type For with probability 1/2 the frequency of A will be 1/3, and with probability 1/2 it will be 1, so the expected frequency of A is $1/2\ (1/3) + 1/2\ (1) = 2/3$ By similar reasoning, the expected frequency of B is $1/2\ (2/3) + 1/2\ (0) = 1/3$ So the A type is fitter, that is, has a higher expected frequency in the second generation, thanks to its lower variance in reproductive output

Another way to put the point is this What matters to evolution is which type will become more *common*, not which type will leave the most offspring These do not necessarily go hand in hand, because how common a type becomes does not depend just on how many offspring it leaves, but also on the total number of offspring produced by all members of the population. In the example above, A and B have the same expected number of offspring, but B's expected frequency after one generation is lower than A's. This is because there is a *positive correlation between the number of offspring* B *leaves and the total number of offspring produced by the population.* If B leaves 0 offspring, then the total number of offspring produced is 5; if B leaves 10 offspring, then the total number produced is 15. So B "dilutes" its own expected frequency, so to speak, by concentrating its offspring in those possible worlds where the total number of offspring produced is large. By contrast A always leaves 5 offspring, so there is no correlation between the number of offspring it leaves and the total number of offspring produced by the population. This is why natural selection favors type A.

The mathematical point here is simple, and has been lucidly explained by Elliott Sober.[16] The frequency of type X, after one generation, is given by

(number of offspring of type X/total number of offspring)

Therefore, the expected frequency of type X, after one generation, is given by

E(number of offspring of type X/total number of offspring)

(We in effect used this formula above to calculate the expected frequencies of A and B after one generation, in the two-member population.) Sober notes, correctly, that the expected value of the quotient of two random variables is *not* in general equal to the quotient of their expected values, that is, $E(X/Y) \neq E(X)/E(Y)$. So from the fact that E(number of offspring of type A) equals E(number of offspring of type B), in our example, it does *not* follow that the expected *frequency* of type A equals the expected *frequency* of type B. This would only follow if it were true that $E(X/Y) = E(X)/E(Y)$, which it is not.

What does this show? The moral most usually drawn, and I think the correct one, is that expected number of offspring is not always the right measure of Darwinian fitness. This may sound odd, for fitness is sometimes *defined* as expected number of offspring. (This is how Skyrms defines fitness in his discussion of the evolutionary Prisoner's Dilemma, for example.) Sometimes this definition is unproblematic, for example, if we assume a large population with independence, or if the variation in reproductive output is the same for every type, but in other cases it is not For the concept of fitness plays a specific theoretical role—the fittest type in a population is the one whose frequency is expected to increase under natural selection If selection is the only evolutionary force at work, fitness differences are meant to predict the evolutionary dynamics Given this theoretical role, it follows that fitness cannot always be equated with expected number of offspring, as the simple example above shows Types *A* and *B* have the same expected number of offspring, in that example, but type *A* is fitter, for its expected frequency after one generation is higher

---

[16] Sober, "The Two Faces of Fitness," in Rama S. Singh, Costas B. Krimbas, Diane B. Paul, and John Beatty, eds., *Thinking about Evolution* (New York: Cambridge, 2001), pp. 309–21.

This leads naturally to the question of how fitness *should* be defined, given that expected number of offspring will not always do the trick[17] Suppose firstly that we relax the infinite population assumption but continue to assume independence A well-known result of John Gillespie,[18] then tells us that the fitness of a type is approximately given by

$$\text{fitness of type } X \approx$$

$$E(\text{number of offspring of type } X) - \{\text{Var (number of offspring of type } X)/N\}$$

where 'Var' denotes the variance function and N is population size In words, this says that the fitness of a type depends not just on its expected reproductive output, but also on its variance in reproductive output, and on the size of the population This formula neatly explains the simple example above Types *A* and *B* have the same expected number of offspring, but *A* has zero variance in offspring number while *B* has positive variance So by the formula above, type *A* comes out fitter Note the role of population size in Gillespie's formula As N gets larger and larger, the variance becomes less and less important in determining fitness and the expectation correspondingly more important, in the infinite limit, variance has no effect on fitness This is highly intuitive Imagine a population containing not two but two *million* organisms—one million *A* and one million *B* In this case *A* has virtually no selective advantage, for the actual reproductive output of the *B*s will deviate only negligibly from its expected value The *A*s will together produce exactly 5 million offspring, and the *B*s will produce 5 million plus or minus a negligible amount (given that each type *B* organism is assumed independent of every other), so virtually no evolutionary change will occur

---

[17] One definition that is always correct, but rarely useful, is simply expected number of offspring relative to the total number in the population, that is, fitness of type $X = E$(number of offspring of type X/total number of offspring) This quantity always predicts the evolutionary dynamics (as A. Grafen shows in "Development of the Price Equation and Natural Selection under Uncertainty," *Proceedings of Royal Society of London B*, CCLXVII, 1449 (2000) 1223–27), but is both difficult to assess empirically and difficult to analyze, given that the denominator depends on the numerator See Frank and Slatkin (*op cit*), and Sober (*op cit*) for further discussion

[18] Gillespie, "Natural Selection for Variances in Offspring Number A New Evolutionary Principle," *American Naturalist*, CXI (1977) 1010–14

This last example also helps illustrates the role of the independence assumption Suppose that instead of independence we assume the extreme opposite perfect correlation, that is, the output of every type *B* organism is the same, in any given generation So however many *B*s there are, they either all produce 10 offspring or all produce zero, each with probability 1/2 (This is sometimes called "temporal" or "between-generation" variance in reproductive output, as opposed to the "spatial" or "within-generation" variance that arises when independence is assumed, see Frank and Slatkin (*op cit*)) In this case, increasing the population size from {1*A*, 1*B*} to {1 million *A*, 1 million *B*} does nothing to reduce the selective advantage of type *A*, for each of the additional *B*s reproduces identically, in each generation More generally, Gillespie (*op cit*) showed that in the presence of perfect correlation among the members of each type, the fitness of a type is approximately given by

$$\text{fitness of type } X \approx$$

$$E(\text{number of offspring of type } X) - \{\text{Var (number of offspring of type } X)/2\}$$

Note the similarity of this formula to the previous formula '*N*' has simply been replaced by 2 This is highly intuitive assuming perfect correlation is effectively equivalent to assuming a population of size N=2, since all the members of a type, however many there are, reproduce identically This shows that relaxing the large population assumption and relaxing the independence assumption are two ways of doing the same thing, both generate a correlation between the total output of the type *B*s and that of the whole population, which constitutes a selective penalty

The foregoing argument shows that risk aversion is evolutionarily advantageous given certain conditions In any small population, or in any population where the independence assumption is violated, there will be selection for reduced variance in reproductive output, that is, for risk-averse behavior How often these conditions are satisfied in nature is an empirical issue that cannot be addressed here, but three points deserve brief mention First, while many evolutionary models assume an infinite or very large population, this is usually for mathematical convenience not because it is a good empirical approximation, second, for much of human evolution, our ancestors are thought to have lived in groups of less than fifty[19], third, the independence assumption does not seem especially plausible either—for it is easy to think of reasons why there may be correlations in reproductive output between the members of a given genotypic or phenotypic class

To summarize, in many realistic biological scenarios, average performance is not the only thing that matters to evolution, variation in performance matters too Natural selection will favor risk-averse organisms over their risk-seeking counterparts, even if both have the same expected reproductive output For the former have lower variance in reproductive output, which is often evolutionarily advantageous So there is a potential Darwinian explanation for the existence of risk-averse behavior

III RISK AVERSION AND THE “PARTING OF WAYS” REVISITED

Recall Skyrms's discussion of the analogy between rational choice and evolution He argues that the analogy is deep, but breaks down in certain cases In the one-shot Prisoner's Dilemma with correlated pairing, evolution and rational choice part ways, because evolution favors a strategy—cooperation—that is at odds with what rational choice theory recommends It is worth looking at the “parting of ways” in relation to the problem of risk aversion

The foregoing analysis demonstrates that given two alternative behaviors, one risk-averse and the other risk-seeking, evolution and rational choice will coincide, *so long as the population is large and the independence assumption is satisfied* Given these conditions, natural selection will not discriminate between the *A* and *B* types, as their expected reproductive output is the same, similarly, the rational agent will not discriminate between options *A* and *B*, as their expected utility is the same But as soon as the large population or the independence assumption is relaxed the situation changes, for expected number of offspring ceases to be a correct measure of fitness, as we have seen The risk-averse *A* type has a fitness advantage, so is favored by selection Evolution and rational choice therefore part ways the preference of the rational agent ceases to coincide with the “preference” of the Darwinian process

This parting of ways is closely analogous to the parting of ways demonstrated by Skryms for the evolutionary PD Skyrms's parting arises from relaxing the assumption of random pairing, ours from relaxing the large population and/or the independence assumptions Interestingly Skyrms[20] himself writes, in relation to the PD, “the foregoing evolutionary model relies on many simplifying assumptions and idealizations that might profitably be questioned", in a footnote he lists these as "random pairing, asexual reproduction, strategies breed true, a large enough population so that we can take expected fitness and average fitness as approximately equal" (*op. cit.*, pp. 118–19n). Skryms explores the consequences of dropping the first of these assumptions, our analysis of dropping the last. In both cases, evolution and rational choice part ways.

---

[19] See R Dunbar, “Coevolution of Neocortex Size, Group Size, and Language in Humans,” *Behavioral and Brain Sciences*, xvi (1993) 681–735

[20] Skyrms, *The Evolution of the Social Contract*, p 53

However, I think that our "parting of ways" is actually more radical than the one shown by Skyrms, for three reasons. First, the issue of random versus correlated pairing arises in a certain specific context, namely the modelling of strategic behavior by game-theoretic methods. Such models are important, but they obviously do not exhaust evolutionary theory. The issue of population size, by contrast, arises everywhere, not just in relation to one specific sort of evolutionary model. Our demonstration that small population size can undermine the link between evolution and rational choice, by inducing selection for risk aversion, is therefore potentially quite significant. It shows that in a wide class of cases, evolution will favor behaviors that conflict with what rational choice theory recommends.

Second, the parting of ways demonstrated by Skyrms in the evolutionary Prisoner's Dilemma rests on the (potentially disputable) assumption of causal decision theory, as discussed previously. In the risk aversion case, by contrast, no such assumption is needed to generate the parting of ways.

Third, our parting of ways puts greater pressure on the analogy between evolution and rational choice than Skyrms's. As we have seen, this analogy rests on the idea that expected utility in rational choice theory plays the same role as expected reproductive output in evolutionary theory, the former determines what the rational agent will choose, the latter what natural selection will choose. Note that this role-isomorphism is retained *even in the one-shot PD with correlated pairing*. In that example, the parting of ways occurs because the strategy with the highest expected utility (defection) is not the same as the strategy with the highest expected reproductive output (cooperation), given CDT, but it is still true that expected reproductive output and expected utility play structurally identical roles—it is just that the occupants of those roles are different. In the risk aversion case, by contrast, expected reproductive output and expected utility do *not* play structurally identical roles. As explained above, in certain situations expected reproductive output is *not* the quantity that matters to evolution, since expected utility is always the quantity that matters to rational choice, the role-isomorphism between these two quantities breaks down.

In our previous example, the expected utility of *A* and *B* is the same—five—as is their expected reproductive output—also five. But rational choice and evolution *still* part ways, because expected reproductive output is not the quantity that determines the evolutionary dynamics. In this respect, our "parting of ways" is qualitatively different to the one explored by Skyrms.

Does this mean that the analogy between evolution and rational choice breaks down altogether, once finite population size is taken into account? I do not think so, though the issue certainly merits further investigation. Rather than pursue this issue further, I want to try to extract a positive moral from the foregoing discussion, concerning how risk-averse behavior in humans might be explained.

IV AN EVOLUTIONARY EXPLANATION FOR RISK AVERSION?

My suggestion is this evolution may have produced in us an inbuilt psychological aversion to risk. The grounds for this hypothesis are as follows. Human psychological tendencies must have been shaped by natural selection, at least in part, evidence from experimental economics shows that humans are pervasively risk averse (though obviously there are exceptions), and there is a general reason for why risk aversion is evolutionarily advantageous, explained above. Obviously this does not *prove* that the hypothesis is true, but it does surely constitute a *prima facie* case, particularly when we bear in mind that rational choice theorists have long struggled to explain the phenomenon of risk aversion.

One feature of this hypothesis deserves comment. The experimental evidence is mainly about *pecuniary* risk aversion, it comes from experiments in which subjects are offered gambles with monetary payoffs. But the evolutionary argument is about gambles with payoffs in offspring, not money, it shows that the expected payoff is not the only thing that matters. Is it legitimate to switch currencies in this way? I think the answer is yes, though care is needed. Many of the things humans desire, including money and material resources, are only indirectly related to reproductive success, at least in modern environments. Nonetheless, it is still reasonable to suggest that evolution has made us averse to gambling with these resources. For over the period in which human psychological tendencies evolved, accumulating material resources, for example, food, must obviously have enhanced survival and reproduction to *some* degree, so risk aversion with respect to such resources would presumably have translated into risk aversion with respect to reproductive output. Thus the "currency shift" is not a problematic feature of our hypothesis, a similar shift is implicit in virtually all theorizing in behavioral ecology.

However, one further, quite subtle issue must be addressed Consider any material resource, for example, food, which enhances reproductive output, or did so in the evolutionary past There are in fact two ways in which risk aversion with respect to this resource might be selected for One is via the evolutionary mechanism described above, that is, selection for reduced variance in reproductive output The second is quite different Reproductive output may simply be a concave function of the resource, that is, further units of the resource bring diminishing gains in reproduction Most resources presumably exhibit such diminishing returns after some level—there is no point having more food than one can eat If so, then risk aversion with respect to the resource will be selected for It would be bad to risk two apples for a one-half chance of gaining four apples, if doubling one's stock of apples less than doubles one's reproductive output Importantly, these two evolutionary mechanisms are logically independent[21] The first works by reducing the variance in reproductive output, the second by increasing the expectation So the first would work even if reproductive output were a linear function of the resource, the second would work even assuming an infinite population with independence Both represent ways in which risk-averse behavior might evolve by natural selection, but the first is more philosophically interesting, as the Skyrms-style "parting of ways" does not occur in the second case

Suppose the hypothesis that evolution has made us risk averse is correct What are the implications for the traditional discussions of risk aversion? In the introductory section, I noted that economists and decision theorists disagree over how pecuniary risk aversion should be explained One camp argues that diminishing marginal utility of money is the explanation, and thus that risk-averse behavior does not violate the expected utility principle at all A second camp disputes this, arguing that the violation of the expected utility principle is real, not apparent the observed extent of risk aversion is more than can plausibly be explained by invoking money's diminishing marginal utility, they argue Human behavior simply does not conform to the prescriptions of expected utility theory, on this view

Which of these camps is right is a complex and much-debated question, here is not the place to attempt a definitive resolution[22] But the evolutionary perspective adopted above may shed some light on the issue If it is true that evolution has made us psychologically averse to risk, this suggests that the second camp may have the better of the argument For given this hypothesis, there is no coherent motivation for trying to explain away our aversion to pecuniary risk by invoking the diminishing marginal utility of money, simply in order to save expected utility theory Better to acknowledge that there is a good evolutionary reason for why we are risk averse, and thus for why we systematically violate expected utility theory From an evolutionary perspective, there is every reason to suspect that the apparent violation of expected utility maximization, evidenced by peoples preference for \$*x* over gambles with an expected payoff of \$*x*, is a real violation

---

[21] Frank and Slatkin (*op cit*) argue that "risk-aversion" should only be used for the first mechanism, the second they call "bet-hedging" However, this terminology is not uniform in the literature

[22] The issue depends in part on interpretation of the empirical evidence and in part on the degree of psychological realism attributed to expected utility theory, for

Another way to put the point is this Theorists in the first camp, who try to explain our aversion to pecuniary risk by invoking money's diminishing marginal utility, are motivated primarily by a desire to preserve the descriptive adequacy of the expected utility model This motivation is a respectable one, for the expected utility model of human decision making has proved its worth in many areas, and in the absence of a reason to think that systematic violations of that model are likely, trying to preserve the model makes sense But evolutionary theory provides just such a reason—it predicts that aversion to risk is likely to be an evolved feature of human psychology, for it is selectively advantageous This implies that systematic violation of expected utility maximization is precisely to be expected, which undermines the main motivation for being in the first camp

A comparison with Skyrms may help illuminate the dialectical situation Skyrms uses the one-shot PD with correlated pairing not just to make the point that evolution and rational choice can part ways, but also for explanatory ends Humans routinely behave cooperatively and altruistically in ways that are hard to explain from a rational choice perspective, as noted previously, experimental subjects often cooperate in the one-shot PD, despite defection being the dominant strategy Skyrms argues that evolution may provide a partial explanation for such behavioral tendencies, so the "parting of ways" has explanatory payoff Similarly, I suggest that risk-averse behavior, which is also very common and yet anomalous from a rational choice perspective, may have an evolutionary explanation Again, the parting of ways is explan-

---

discussion of the former, see Rabin (*op cit*), for the latter, see Bengt Hansson, "Risk Aversion as a Problem of Conjoint Measurement," in Peter Gärdenfors and Nils-Eric Sahlin, eds , *Decision, Probability and Utility* (New York Cambridge, 1988), pp 136–58 atorily useful: it suggests an explanation for why our behavior systematically violates expected utility theory.

Note that in both Skyrms's case and our own, defenders of expected utility theory may *try* to argue that the behaviors in question—cooperation and risk aversion respectively—do not really violate the theory at all. In the risk-aversion case, they can appeal to diminishing marginal utility of money, as discussed; in the one-shot Prisoner's Dilemma, they can argue that in fact, humans care not just about their own payoff but also about the payoff of their partner, so cooperating is the rational strategy.[23] This illustrates the well-known point that virtually any behavior can be rendered compatible with expected utility theory by making ad hoc postulations about the agent's utility function. Methodologically this is a suspicious move, for it threatens to deprive expected utility theory of any predictive content, but many theorists have nonetheless been tempted by it. In the light of evolutionary considerations, the temptation should seem much less attractive. For evolution supplies a principled reason for *why* behaviors such as cooperation and risk aversion are so prevalent in humans, thus removing the motivation for trying to force such behaviors into the expected utility framework by ad hoc means.

Finally, let us consider normative issues. If our aversion to risk is partly the result of evolution, what does this imply about the rationality of risk-averse behavior? Does it mean that evolution has programmed us to be irrational? This is a tricky question. The answer partly depends on what we think of the normative status of expected utility theory. The standard argument for saying that rationality requires conformity to the expected utility principle comes from the representation theorems of Leonard J. Savage,[24] Richard Jeffrey,[25] and others—which show that unless we choose the option with the highest expected utility, our preferences violate the axioms of decision theory. However, other theorists, for example, Edward McClennen,[26] dispute the force of this argument, essentially because they do not regard those axioms as self-evident rationality requirements. Skyrms himself is cautious about claiming that evolution has led to irrationality, despite his "parting of ways" argument, and I think his caution is well-placed.

---

[23] This assumes of course that payoffs are measured in units of money, or some other resource, not utility. If payoffs were measured in utility, the argument would simply be that the payoff matrix characteristic of the Prisoner's Dilemma describes a situation in which humans almost never find themselves.

[24] Savage, *The Foundations of Statistics* (New York: Wiley, 1954).

[25] Jeffrey, *The Logic of Decision* (Chicago: University Press, 1965).

[26] McClennan, "Sure-Thing Doubts," in Gärdenfors and Sahlin, eds., pp. 166–82.

Not only is it difficult to assess whether apparent violations of expected utility maximization are real, it is also difficult to assess the normative claim that such violations are irrational So even if my hypothesis that evolution has led us to be risk averse is correct, it does not automatically follow that evolution is guilty of making us irrational

CONCLUSION

The arguments presented here do not pretend to be conclusive But so long as we do not reject out of hand the idea that evolutionary reasoning can help illuminate human psychology, the hypothesis that risk aversion is a Darwinian adaptation deserves to be taken seriously Additionally, this provides further evidence of the insights that can be gained by adopting an evolutionary perspective on problems traditionally discussed within a pure rational choice framework

SAMIR OKASHA

University of Bristol

THE LONG ROAD TO SKEPTICISM*

The ancient skeptic had two different roads to skepticism. The *short* road is embodied most clearly in the Five Modes of Agrippa, which offer a perfectly general way to provoke a dogmatist into an inquiry he cannot complete, even to his own satisfaction. (The Ten Modes of Pyrrhonism are a somewhat more scenic version of the short road.) The *long* road, by contrast, is not general at all. Rather, it is quite specifically focused on the leading bodies of knowledge and theories of knowledge of the day.[1]

In modern skepticism the road most followed has been the short one. Consideration of skepticism has been driven by the various scenarios—evil demons, brains in vats, virtual reality matrices—that so well capture our pre-theoretic imagination. Little if any attention, however, has been given to the long road. And that is a shame, for a great deal of interesting epistemology came out of the ancient long road. In what follows, then, I shall take a few steps down the long road, if only to see where it leads now.

I

In following the ancient long road, Sextus gives us arguments against (among other things) physics, mathematics, and linguistics, as well as some against the epistemological views of his day. Ours is a less patient age, however, and so I shall take a shortcut. I will restrict my focus merely to the epistemological views of our day. Here, I am helped by the fact that, despite the diversity of views currently on offer, there are two fundamental theses to which all of them are committed and that also, as I shall argue, leave them vulnerable to skeptical objections.

The first thesis is *fallibilism*, which we may think of as being simply the combination of two highly plausible claims: (1) on many occasions we do in fact have knowledge, and yet (2) given the fallible

---

* For comments on earlier versions of this paper, I am grateful to Michael Huemer, Jason Kawall, Jaegwon Kim, Ernest Sosa, and Jim Van Cleve. I owe a special debt of gratitude to Jennifer Lackey—for her comments on every version of this paper, for her discussions with me of everything epistemological, and for so much else.

[1] For the classic version of the short road, see Sextus Empiricus's *Outlines of Pyrrhonism*, he distinguishes the two different approaches to skepticism at i.2. Examples of the long road include both Sextus's *Against the Mathematicians* and Cicero's *Academica*.

0022-362X/07/0405/236–62

nature of our cognitive faculties we could have failed to have had knowledge on those occasions More precisely

> FALLIBILISM For any proposition *p* that *S* in fact knows on the basis of justification *j*, *S*'s belief that *p* could have been either (i) false or (ii) accidentally true, even while being held on the basis of *j*[2]

The justification that grounds the subject's belief as knowledge does so because it makes that belief *probable*[3] Having justification for a belief thus does not guarantee that the belief will be nonaccidentally true [4] My visual experience of a tree, for example, does not guarantee that there really is a tree there I could, after all, be suffering from a hallucination Still, it is hard to deny that the experience is an excellent basis for believing that the tree is there If we are to have knowledge without guarantees, then, the world must "do us a favor," as John McDowell puts it [5] Fallibilists disagree as to what that favor amounts to, but none of them reject the favor itself As Stewart Cohen says, "the acceptance of fallibilism in epistemology is virtually universal "[6]

II

The second fundamental thesis, which we may call *attributabilism*, requires a bit more explication It has frequently been accepted in an explicit way—though usually in a logically stronger form—by a wide range of philosophers, and, as will become clear, it has long functioned as a deep, implicit commitment in epistemology

---

[2] See my "How to Think about Fallibilism," *Philosophical Studies*, CVII (2002) 143–57, for this characterization of fallibilism For the sake of simplicity, I have formulated fallibilism as a universal claim Many philosophers, however, would be willing to allow that at least a small number of propositions (including, for example, the *cogito*) can be known infallibly Knowledge of this sort, if any is possible, will not be affected by the argument that follows I should also note that I am using the term 'justification' in a broad sense, so that it may cover a wide range of epistemological theories, thus including both externalist theories like reliabilism (in which probability is explicitly one of the central elements) and internalist theories like moderate foundationalism and coherentism (which allow for a belief to be justified even though it is not certain), for reliabilism, see, for example, Alvin Goldman, *Epistemology and Cognition* (Cambridge Harvard, 1986), for foundationalism, see Roderick Chisholm, *Theory of Knowledge* (Englewood Cliffs, NJ Prentice Hall, 1989, 3rd edition), and for coherentism, see Laurence BonJour, *The Structure of Empirical Knowledge* (Cambridge Harvard, 1985) The only restriction is that justification, as understood here, is a necessary condition for knowledge rather than, say, merely rational acceptability

[3] See my "How to Think about Fallibilism "

[4] A belief is accidentally true (as, for example, in a typical Gettier case) when the explanation for why the belief is held is independent of what makes the belief true See my "Accidental Truth and Accidental Justification," *The Philosophical Quarterly*, L (2000) 57–67, for more on accidentality, and Edmund Gettier, "Is Justified True Belief Knowledge?" *Analysis*, XXIII (1963) 121–23, for the original Gettier cases

[5] McDowell, "Knowledge and the Internal," *Philosophy and Phenomenological Research*, LV (1995) 877–93

[6] Cohen, "How to Be a Fallibilist," *Philosophical Perspectives*, II (1998) 91–123, on p 91 Michael Williams agrees "We are all fallibilists nowadays", see his "Contextualism, Externalism and Epistemic Standards," *Philosophical Studies*, CIII (2001) 1–23, on p 5

Let us begin with a rough statement of the thesis

> ATTRIBUTABILISM For any state of knowledge *k* that a person *S* possesses, the possession of *k* is attributable to *S*

At first glance, it looks like a truism—if *S* knows that *p*, then of course the knowledge that *p* is attributable to *S* But what is meant by "attributable" is something stronger A comparison with the attributability of actions is instructive here

There are different ways of explaining the special relationship an agent bears to her action According to simple versions of compatibilism, like those of Hobbes and Hume, the action must be caused by the agent's will For more sophisticated versions of compatibilism, the action must come from what Susan Wolf calls the "Real Self," where some philosophers take this to be an appropriately structured will while others argue it is the agent's faculty of reason [7] Among libertarians, who attempt to ground responsibility for actions in what Robert Kane calls "ultimate responsibility," we find similar variety some of them require only that the action be indeterministically produced by the agent (perhaps on the basis of reasons the agent has), while others hold the agent to be a special sort of cause [8] But what is common to all of these philosophers, compatibilists and libertarians alike, is the attempt to show how the action came from the *person* in question [9] It is not enough for the person to be merely the locus of a causal process terminating in something like an action [10] A reflexive leg-kicking, for example, is not something that one *does*—it fails to belong, in the relevant sense, to the person who kicks

---

[7] Wolf, *Freedom within Reason* (New York Oxford, 1990), chapter 2 See Harry Frankfurt, "Freedom of the Will and the Concept of a Person," this JOURNAL, LXVIII, 1 (January 14, 1971) 5–20, for the view that freedom and responsibility stem from a hierarchically structured will, and Gary Watson, "Free Agency," this JOURNAL, LXXII, 8 (April 24, 1975) 205–20, for the view that they result from reason Though Wolf distinguishes her own view from Watson's (*Freedom within Reason,* p 30), they are in many ways similar The simpler version of compatibilism mentioned above may also be a version of the "Real Self" view if, as seems plausible, those philosophers took the real self to be nothing more than the will

[8] See Robert Kane, *The Significance of Free Will* (New York Oxford, 1996), pp 33–35, on ultimate responsibility Kane advocates a view that falls into the first broad category, see Chisholm, "Human Freedom and the Self," The Lindley Lecture at the University of Kansas, 1964 (reprinted in Watson, ed , *Free Will,* (New York Oxford, 1982), pp 24–35), for an example of the latter sort of account

[9] In fact, Frankfurt (*op cit*) explicitly links freedom of the will with the concept of a person

[10] On this point, see J David Velleman, "What Happens When Someone Acts?" *Mind,* CI (1992) 461–81

Many epistemologists have made a similar point about knowledge and its relation to the person who has it, though often in service of rather different projects For example, in attempting to explain the incompatibility of knowledge with epistemic luck, Duncan Pritchard says that "Knowledge does appear to be a cognitive achievement of some sort Accordingly, to say that an agent has knowledge is to imply that she didn't gain that knowledge *simply* via good fortune "[11] In arguing against a safety requirement on knowledge, Ram Neta and Guy Rohrbaugh claim that "Knowledge is an important cognitive achievement Like other achievements worth pursuing, it must be earned and is not assured "[12] And Ernest Sosa, Wayne Riggs, and John Greco argue that conceiving of knowledge as an achievement—something for which we deserve credit—allows for an explanation of the special value of knowledge Sosa, for example, says that "*What we prefer is the deed of true believing, where not only the believing but also its truth is attributable to the agent as his or her own doing*"[13] In addition, Riggs and Greco use the claim that knowledge is something for which we deserve credit to distinguish cases of knowledge from cases of accidentally true beliefs (that is, Gettier cases) So, Riggs holds that "Being in the state of 'knowing that *p*' entails of a person that she have a true belief for which she deserves a certain degree of epistemic credit Believing something true by accident entails no credit of any sort to the person "[14] Similarly, Linda Zagzebski links knowledge with the person's having a desire for the truth "My getting the truth [without having a desire for the truth] is a matter of luck in a certain sense of luck I don't get credit for getting the truth I suspect that this is sufficient to deprive me of knowing "[15]

---

[11] Pritchard, *Epistemic Luck* (New York Oxford, 2005), p 1

[12] Neta and Rohrbaugh, "Luminosity and the Safety of Knowledge," *Pacific Philosophical Quarterly*, LXXV (2004) 396–406, on p 404 The safety requirement they reject is endorsed by Ernest Sosa, "How Must Knowledge Be Related Modally to What Is Known?" *Philosophical Topics*, XXVI (1999) 373–84, and by Timothy Williamson, *Knowledge and Its Limits* (New York Oxford, 2000)

[13] Sosa, "The Place of Truth in Epistemology," in M DePaul and L Zagzebski, eds , *Intellectual Virtue Perspectives from Ethics and Epistemology* (New York Oxford, 2003), pp 155–79, on p 175 (emphasis in the original) See also Riggs's "Beyond Truth and Falsehood The *Real* Value of Knowing that *P*," *Philosophical Studies*, CVII (2002) 87–108, and "Reliability and the Value of Knowledge," *Philosophy and Phenomenological Research*, LXIV (2002) 79–96, and Greco's "Knowledge as Credit for True Belief," in DePaul and Zagzebski, eds (*op cit*), pp 111–34

[14] Riggs, "Reliability and the Value of Knowledge," p 94 See also Greco (*op cit*)

[15] Zagzebski, "Intellectual Motivation and the Good of Truth," in DePaul and Zagzebski, eds (*op cit*), pp 135–54, on p 151

Now, just as we saw in the case of action, there are many different accounts of the special relationship a person has to her knowledge—that is, there are different ways of understanding what sort of an "achievement" knowledge is. For example, some epistemologists take knowledge to be grounded in epistemic duties. Thus, Roderick Chisholm construes his basic epistemic locution, "A is more justified than B for S," as expressing a requirement to prefer.[16] It is clear, of course, that any such requirement could apply only to a person. A subpersonal belief-forming process is not capable of fulfilling epistemic duties or meeting requirements to prefer.

A view like Chisholm's is often thought to lend support to the claim that we are *responsible* for our knowledge, where this entails that the subject has a direct control over what she believes. But it is important to carefully distinguish attributabilism from this stronger, and far more contentious, claim. *Responsibilism*, as we may call it, entails attributabilism, but not vice versa. Serious doubts have been raised regarding whether we do in fact have so strong a form of control as that.[17] But that debate does not concern us, for as we shall see there are many philosophers who reject responsibilism but nevertheless accept attributabilism.

Now, it may be obvious that classic internalists like Chisholm and Laurence BonJour, who take knowledge to be intimately connected with one's subjective perspective, are committed to attributabilism.[18] It is also clear that some versions of externalism—like Greco's agent reliabilism and Sosa's virtue perspectivism—are explicitly committed to attributabilism. Nevertheless, one might wonder how externalism could be compatible with this thesis. Externalists, after all, take knowledge to be grounded in some relation between the subject and the world—for example, reliability—that typically falls outside the ken of the subject. And if that is true, it is difficult to see how the reliability of her belief-producing faculties can be within the subject's control. Furthermore, one might argue that reliability cannot be attributed to the subject if it is not within her control. But if the reliability of her belief-forming faculties is not within her control, then surely the subject cannot be said to have control over whether the beliefs produced by those faculties are reliable. The concern, then, is that on an externalist view the subject seems not to have control over whether her beliefs are knowledge, and for that reason it looks as though knowledge is not genuinely attributable to the subject.

---

[16] Chisholm, *Theory of Knowledge*, p. 59.

[17] See, for example, William Alston, "The Deontological Conception of Epistemic Justification," *Philosophical Perspectives*, II (1988): 257–99. For an argument that we have an *indirect* control over what we know, grounded in the direct control we have over what we do, and that this is sufficient to make epistemic agency possible, see my "Epistemic Agency and the Intellectual Virtues," *The Southern Journal of Philosophy*, XXXIX (2001): 507–26. I argue there, however, that epistemic concepts like *justification* and *knowledge* should not be analyzed in terms of epistemic agency concepts like *duty* or *virtue*.

[18] See BonJour, *The Structure of Empirical Knowledge*, p. 8: "My contention here is that the idea of avoiding such [epistemic] irresponsibility, of being epistemically responsible in one's believings, is the core of the notion of epistemic justification."

This worry is grounded in the purported link between attributability and control. Here, it is again instructive to consider the attributability of actions. It is quite clear that libertarians *do* accept the link between attributability and control. The agent bears ultimate responsibility for her actions because it is she—and she alone—who determines which action she will take. But it is also clear that compatibilists reject or significantly qualify the link between attributability and control. Compatibilists generally find the idea of ultimate responsibility to be mysterious and argue that the agent has been shaped by the world in which she finds herself. What matters is not the cause of her practical identity but rather that her actions are caused by her practical identity.[19] For compatibilists, then, there are limits to how much control an agent can have over her action. She may modify the relevant aspects of her practical identity—for example, she may attempt to acquire virtues or to break bad habits—but she does so only because this action is itself caused by the combination of her given practical identity and her circumstances.

Now, returning to the attributability of beliefs, we can see that externalists are in a position to similarly reject or qualify the link between attributability and control. Whether or not reliability is outside of the subject's ken or control is irrelevant. What matters is that we use the faculties, shaped by the world to be reliable, that we have. Our basic reliable faculties constitute who we are cognitively, prior to any choices we make about who we want to be; in this sense, our cognitive characters are *world-involving*.[20] We do not need to ensure that our faculties are reliable prior to using them—it is enough that they *are* reliable.[21] On this view, then, knowledge is attributable to us, not because we have ultimate control over it, but because it comes out of a cognitive "Real Self."

---

[19] Cf. Frankfurt, *The Importance of What We Care About* (New York: Cambridge, 1988), p. viii: "Each of us is indisputably a creature of history. Without the past, we should be nothing, and it is only by appreciating that he is through and through the product of historical contingencies—biological, social, and personal—that a person can recognize and make sense of his own nature."

[20] There is another way in which our cognitive characters are world-involving. The reliability of a faculty is always relative to a subject matter and to a type of environment. Vision, for example, is reliable relative to normal conditions in the actual world, but it would not be reliable in a world where the atmosphere distorted light in unpredictable ways. It is impossible, then, to evaluate our faculties apart from the environment in which they are functioning; we do not have any cognitive character in isolation from the rest of the world. I shall say more about this below.

Something like this is central to virtue epistemology, of both the Aristotelian and the reliabilist kinds. Both varieties of virtue epistemology take epistemic excellence to be grounded in the nature of the cognitive subject as a whole.[22] Virtues are ways in which a *person* is performing successfully; they do not have their value independently of our evaluation of the person who has them. Thus, Greco says that "When we attribute knowledge to someone we imply that it is to his credit that he got things right. It is not because the person is lucky that he believes the truth—it is because of his own cognitive abilities."[23] Similarly, Sosa holds that

> We prefer truth whose presence is the work of our intellect, truth that derives from our own virtuous performance. We do not want just truth that is given to us by happenstance, or by some alien agency, where we are given a belief that hits the mark of truth not through our own performance, not through any accomplishment creditable to us.[24]

These virtue epistemologists also hold a stronger thesis than attributabilism. Sosa says that what is most important "is your grasping the truth attributably to your intellectual virtues acting in concert conducted by reason, and thus attributably to you as epistemic agent ”[25] In saying so, he is accepting not only the thesis of attributabilism but also advancing a particular view about what conditions are necessary for the proper attribution of knowledge to the subject This latter view may be criticized by some epistemologists—not everyone thinks that knowledge must derive from intellectual virtues guided by reason—but, again, the more basic claim of attributabilism appears to be shared universally

---

[21] For more on this distinction, see James Van Cleve, "Foundationalism, Epistemic Principles, and the Cartesian Circle," *Philosophical Review*, LXXXVIII (1979): 55–91. Of course, once we have begun using our reliable faculties, there is a great deal we can do to shape our own epistemic development; on this point, see Van Cleve (*op. cit.*) and Sosa, "How to Resolve the Pyrrhonian Problematic: A Lesson from Descartes," *Philosophical Studies*, LXXXV (1997): 229–49.

[22] For example, Sosa writes, "Virtue epistemology is distinguished by its emphasis on the subject as seat of justification"—"Reflective Knowledge in the Best Circles," this JOURNAL, XCIV, 8 (August 1997): 410–30, see p. 419. See also Sosa, "The Place of Truth in Epistemology"; Zagzebski, *Virtues of the Mind* (New York: Cambridge, 1996), and "Intellectual Motivation and the Good of Truth"; Riggs, "Reliability and the Value of Knowledge"; and Greco, *Putting Skeptics in Their Place* (New York: Cambridge, 2000), and "Knowledge as Credit for True Belief."

[23] Greco, "Knowledge as Credit for True Belief," p. 123. See also Riggs's "Beyond Truth and Falsehood," and "Reliability and the Value of Knowledge."

[24] Sosa, "The Place of Truth in Epistemology," p. 174. In "Why We Don't Deserve Credit for Everything We Know," *Synthese* (forthcoming), Jennifer Lackey objects that in some cases of genuine knowledge—for example, getting directions from a stranger—we do not deserve credit for obtaining the truth. Rather, the credit goes to the person who gave the directions to us. (Lackey also presents cases of knowledge where the credit does not go to another person but to some environmental feature.) Although cases of this sort pose a problem for the credit view espoused by Greco and Sosa, they are not incompatible with attributabilism. Just as an action (for example, buying a valuable antique) may be attributable to a person who does not deserve credit for it (the person's wife is the one who found it), so an instance of knowledge—for example, in the case Lackey mentions—may be attributable to a person who does not deserve credit for it.

This is evident from the typical response externalists of every stripe have given to BonJour's counterexamples, in which a subject forms a true belief using what is in fact a reliable belief-producing process but also has good counterevidence for the belief In one such case

> Maud believes herself to have the power of clairvoyance, though she has no reasons for this belief She maintains her belief despite being inundated by her embarrassed friends and relatives with massive quantities of apparently cogent scientific evidence that no such power is possible One day Maud comes to believe, for no apparent reason, that the President is in New York City, and she maintains this belief despite the lack of any independent evidence, appealing to her alleged clairvoyant power Now in fact the President is in New York City, and Maud does, under the conditions then satisfied, have completely reliable clairvoyant power Moreover, her belief about the President did result from the operation of that power [26]

Two responses are possible for the externalist She may say either (1) that the subject's belief counts as knowledge, despite the irrationality of holding a belief in the face of good counterevidence, or (2) that the irrationality is unacceptable and the counterevidence therefore provides a *defeater* so that the belief does not count as knowledge By and large, externalists have given the second response [27] But notice that to do so just *is* to accept attributabilism In adding the no-defeater condition to their views, externalists are in effect saying that what truly matters epistemically is not just the relation between the

---

[25] Sosa, "The Place of Truth in Epistemology," pp 178–79 Similarly, Greco advances an account of intellectual credit where this is taken to be achieved only when a subject's true belief "reveals" the subject's reliable cognitive character, see his "Knowledge as Credit for True Belief," p 123

[26] BonJour (*op cit*, p 40), this is the third of seven cases

[27] See, for example, Robert Nozick, *Philosophical Explanations* (Cambridge Harvard, 1981), Goldman (*op cit*), and Alvin Plantinga, *Warrant and Proper Function* (New York Oxford, 1993) For more on defeaters, see Lackey, "Testimonial Knowledge and Transmission," *The Philosophical Quarterly*, XLIX (1999) 471–90, and my "Epistemic Circularity Squared? Skepticism about Common Sense," *Philosophy and Phenomenological Research*, LXXIII (2006) 186–97

belief and the world, or even the relation between the relevant belief-producing *process* and the world, but rather the broader relation between the *person* and the world, of which the belief and the belief-producing process are only parts. The reliability of the true belief is not itself enough for knowledge simply because this excellence, when it occurs in the context of irrationality, is not attributable to the person.

To fully appreciate this, notice that it makes no sense to say that the counterevidence shows the belief itself to be problematic independently of what is happening in the person as an epistemic whole. It is not, after all, the *belief* that disregards the available counterevidence—only the *person* can do that. In fact, the only way in which we can meaningfully say that there *is* counterevidence available in this situation—and that the counterevidence is therefore in a position to render the subject irrational in holding a given belief—is by reference to the epistemic capacities of the person in question. If, for example, the subject in one of BonJour's cases was informed in Swedish that clairvoyance was unreliable, and yet she does not understand a word of Swedish, we would not take her continued use of clairvoyance to be irrational.[28] This information would be, in a sense, invisible to her and thus would not count against the rationality of her belief. So, given that the epistemic defect in the original BonJour cases derives, not from the belief considered alone, but from the epistemic condition of the person who has it, it is clear that a state of knowledge—a situation free from epistemic defect—would likewise be grounded in the performance of the person, and not merely in the excellence of the belief.

To *deny* attributabilism would be to accept some form of what Wilfrid Sellars denigrated as a "thermometer model" of knowledge.[29] To be sure, what a thermometer does is in some way analogous to knowledge, and perhaps we may fruitfully use that analogy to shed some light on what it is. But if we push the analogy too far, we lose sight of the proper focus of the normative aspect of knowledge. In the sort of evaluation we perform when we inquire as to whether someone knows something, we look, not at subpersonal belief mechanisms or at isolated beliefs, but at how good the overall response on the part of the person is—we look, not merely at what actually happens, but at what *should* happen. This is a concern that is recognizable only at the level of the *person*—not at the subpersonal level.

---

[28] For more on this point, see Lackey, "Testimony and the Infant/Child Objection," *Philosophical Studies*, cxxvi (2005): 163–90.

[29] Sellars, "Empiricism and the Philosophy of Mind," in *Science, Perception, and Reality* (Atascadero, CA: Ridgeview, 1963), pp. 127–96.

The problem with the thermometer model is not its mechanistic character but the way it encourages a fragmentary view of cognition Suppose there were a mechanical weather forecaster, which used a variety of instruments, including a thermometer On one occasion, the thermometer, functioning reliably, causes the forecaster to predict sunny weather (and in fact this is what eventually transpires) However, the forecaster is malfunctioning and "ignores" the conflicting information that some of its other instruments are providing Does it "know" that sunny weather is on the way? No—it *should not* be providing this forecast, given that malfunction Again, it is the assessment of the cognitive system that matters here There is no sense in which the thermometer (rather than the weather forecaster) should be doing something different than it does Similarly, in BonJour's cases there is no sense in which the subject's faculty of clairvoyance should be doing something different *Its* functioning is perfectly fine The problem is with the *person* who has the clairvoyance

It should now be clear, I hope, how deep and widespread the commitment to attributabilism is We could not eliminate it from our conception of knowledge without thereby fundamentally changing it What would be left over would not be knowledge as we now think of it but rather something instead like *possessing information* And that concept, although it might be useful for some purposes, would not at all be an appropriate substitute for many of the contexts in which we now require a subject to have knowledge [30]

III

Here is where the long road has led us There are two theses about knowledge that are of central importance to the epistemological theories of our day, fallibilism and attributabilism But, as I shall now argue, these are not compatible with one another If fallibilism is true, then attributabilism is not Thus, knowledge, as understood by the prevailing theories, is impossible In that sense, then, we do not know anything

So, let us turn now to a defense of the conditional claim that, if fallibilism is true, then attributabilism is not One way to think about the basic epistemological project is to see it as the attempt to explain what it means for someone to have a true belief where the truth of the

---

[30] For example, it is plausible to suppose that knowledge rationalizes action But it is possible to conclude that an action is the on the whole best thing to do only if it optimally fits the agent's cognitive state This will very often be impossible to determine if we evaluate the agent's cognitive state just in terms of the information it contains, given that someone who has evidence on both sides of a particular question will be in informational states that conflict with one another

Colombo University Library

belief is no accident, as happens, for example, when one makes a lucky guess. An important first step is to impose a *justification* condition on knowledge.[31] A justified belief differs from a lucky guess in that the justified beliefs are likely to be true, whereas guesses are not.

As Edmund Gettier (*op. cit.*) famously showed, however, justified true belief is not sufficient for knowledge, so long as we are working within the framework of fallibilism. Recall that, according to fallibilists, the justification that in fact grounds a subject's belief as knowledge does not guarantee that the belief will be knowledge. It is possible that the subject could have held the belief with that very same justification and yet it could have failed to be knowledge in two ways: it could have been false, or it could have been true, but only accidentally so. For example, although the subject actually saw a barn, she could have had the same experience while instead seeing a mere barn façade, or she could have had the same experience while seeing the only genuine barn in a county otherwise filled with mere barn façades.[32] The basic epistemological project has not yet been carried out. We do not yet have an explanation of what it means to have a true belief where the truth of the belief is no accident.

Some fourth condition on knowledge, in addition to truth, belief, and justification, is apparently required.[33] (Let us call this condition *c*.)

---

[31] I am here continuing to use the term 'justification' to refer to one of the necessary conditions for knowledge rather than mere rational acceptability.

[32] The latter type of case appears in Goldman, "Discrimination and Perceptual Knowledge," this JOURNAL, LXXIII, 20 (November 18, 1976): 771–91, and is due to Carl Ginet.

[33] Williamson argues that the traditional approach in epistemology, which attempts to analyze knowledge through the provision of necessary and sufficient conditions (for example, truth, belief, justification, and so on), fails because it cannot solve the Gettier problem (*op. cit.*, pp. 8–9, 30–33). In its place, he offers a "non-conjunctive" view, according to which knowledge is the most general kind of factive mental state. If Williamson's view could avoid the Gettier problem, it would not face the argument for skepticism I am developing here. However, as I argue in "Accidentally Factive Mental States," *Philosophy and Phenomenological Research*, LXXI (2005): 134–42, the Gettier problem can be reformulated so that it is just as problematic for Williamson's view as it is for traditional epistemology. And so, the argument for skepticism presented here could similarly be reformulated so as to apply to a view like Williamson's. The same holds for other accounts of knowledge that do not make use of a justification condition—see, for example, Plantinga's account of warrant in terms of proper function (*op. cit.*). (Plantinga also characterizes warrant as what makes the difference between knowledge and mere true belief; see his *Warrant: The Current Debate* (New York: Oxford, 1993). See also Trenton Merricks, "Warrant Entails Truth," *Philosophy and Phenomenological Research*, LV (1995): 841–55. There cannot be accidentally true beliefs that are warranted, in this sense of *warrant*. But this does not mean that the Gettier problem has been resolved. Rather, it means that warrant, in this sense, cannot be accounted for by Plantinga's proper function theory; on this point, see Richard Greene and N.A. Balmert, "Two Notions of Warrant and Plantinga's Solution to the Gettier Problem," *Analysis*, LVII (1997): 132–39. The skeptical argument presented above, then, can be taken as showing that warrant—understood as whatever converts true belief to knowledge—is impossible.)

But here is where the conflict with attributabilism arises Whatever condition *c* turns out to be, it cannot be something that is attributable to the person in question This is evident for two reasons First, the justification condition *already* represents the full extent of the subject's contribution toward having a nonaccidentally true belief We cannot simply require that the subject do *more* toward having a nonaccidentally true belief, for, if fallibilism is true, there will always be a gap between what the subject does and the non-accidental truth of the belief We cannot get rid of that gap without also getting rid of fallibilism Second, reflection on a typical Gettier-type case shows that the problematic accidentality arises from environmental factors that are completely extrinsic to the subject In the barn façade case mentioned above, it is the presence of barn façades in the vicinity of the subject that makes her belief accidentally true Condition *c*—whatever formulation it takes—would have to preclude the nearby presence of those barn façades This simply has nothing to do with anything that we may properly attribute to the subject The presence or absence of the barn façades makes no difference at all to the nature of the subject's contribution toward having a nonaccidentally true belief

Notice that this will be so even for externalist theories that take the cognitive nature of the person to be world-involving Suppose, for example, that a person's cognitive character is constituted by reliable belief-producing processes or virtues Given that reliability supervenes, not only on the subject's belief-forming processes but on general facts about the surrounding environment as well, we will have to grant that the subject is constituted in part by that environment Those general facts are external to the person's subjective awareness, but they are *not* extrinsic to the subject's cognitive nature However, the same cannot be said about the environmental factors whose presence may give rise to problematic accidentality These are different in kind from the much more general features of the surrounding environment that help constitute reliability and are therefore intrinsic to the subject[34]

---

[34] Thus, when Plantinga responds to the Gettier problem by requiring that the subject be in an appropriate "cognitive mini-environment" (in "Warrant and Accidentally True Belief," *Analysis*, LVII (1997) 140–45), he is not merely extending the reasoning underlying his initial account of warrant, which holds that "your faculties must be appropriate for your particular repertoire of epistemic powers It must be the sort of environment for which your faculties are designed—by God or evolution (or both)" (*Warrant and Proper Function*, p 7) Surely, one's cognitive character has not been designed—by either God or evolution—so as to take into account the absence of fake barns in a particular locale and on a particular afternoon Adding the mini-environment condition, then, is a substantive change to Plantinga's account It also leads to the following sort of problem if we take it to be a necessary condition of warrant that the

So, when the factors that might lead to accidentality are absent, this is not something that can be attributed to the subject

The result, then, is that the person's cognitive performance is not good enough to allow her to reach the truth Her performance must be supplemented by something extrinsic—and therefore unattributable—to her Thus, knowledge itself is not attributable to the person, either Attributabilism is therefore false, as long as we are committed to fallibilism *A fortiori*, any theory of knowledge committed to attributabilism—which is to say, *every* theory of knowledge—is also false [35]

IV

What are the possible responses to this type of skepticism? There are only two options, neither is attractive First, epistemologists might try rejecting fallibilism But, of the various forms of infallibilism that have been developed in the philosophical tradition (for example, by the Stoics, by Descartes, by some of the twentieth-century sense-datum theorists), none are now seen as being live options That stretch of the long road to skepticism has already been traveled

The other option is to reject attributabilism There are three arguments for doing so First, truth is generally taken to be a necessary condition of knowledge, yet it is rarely attributable to the person who purportedly has knowledge So, it looks as though knowledge itself cannot be attributable to the person, either [36] Second, epistemologists have known for a long time that some condition in addition to justification is needed to rule out cases of accidentally true belief The fact that this nonaccidentality (or "anti-Gettier case") condition conflicts with attributabilism merely shows that epistemologists do not actually accept that thesis [37] Third, even if we reject attributabilism, which links attributability with knowledge, we are free to accept a closely related, but weaker thesis for any justified belief *j* that a person

---

subject be in a favorable mini-environment, it will turn out to be a matter of luck that anyone ever has a warranted belief After all, nothing the subject does determines whether she is in the right or the wrong sort of mini-environment, relative to her cognitive performance, it will thus be an accident when she is in the right one

[35] It should be noted that the argument presented above is not a mere restatement of the Gettier problem Gettier (*op cit*) provided counterexamples to some prominent analyses of knowledge, which leaves open the possibility that the problem posed by those counterexamples can be solved The argument presented in this paper denies this possibility The conclusion is not simply that no one *has* solved the problem of accidentality but rather that the problem *cannot* be solved (without making radical changes to our current shared conception of knowledge)

[36] I am grateful to Jaegwon Kim for this objection

[37] I am grateful to Michael Huemer for this objection

S possesses, the possession of *j* is attributable to *S*. This proposal captures everything that made attributabilism initially seem plausible, yet it avoids the problems mentioned in the first two arguments.[38]

Let me begin with the first objection by returning to the attributability of actions—for example, climbing a mountain. This is something that can be fully attributable to a climber (assuming nothing unusual—for example, being carried part of the way by a sherpa guide—happens). Nevertheless, in order to climb a mountain there must be a mountain, a fact that is surely not attributable to the average climber. This does not, however, undermine the attributability of the climber's achievement. The existence of the mountain provides the goal at which she aims. *Given* that the mountain is there, her reaching the top of it is something that can be entirely her own doing. Similarly, in the case of knowledge, the sense in which the truth of a belief is not attributable to the person who has it is irrelevant epistemically. In order to know that it is raining outside, I do not have to make it rain. *Given* that it is raining, however, I may come to know it. The truth of the proposition provides the goal at which I aim—whether or not I reach it is up to me.

There are two points to note in response to the second objection. First, although epistemologists have long recognized the need to rule out Gettier cases, many have tried to do so by adapting an externalist account of knowledge to this end—for example, Alvin Goldman's relevant alternatives approach, Robert Nozick's tracking account, Alvin Plantinga's proper functionalism, Greco's agent reliabilism, and Pritchard's safety-based account.[39] Thinking of the problem in this way, however, has blurred the difference between factors that are *extrinsic* to the subject's cognitive performance and factors that are merely *external* to the subject's awareness. The latter—which feature in the views in question—are compatible with attributabilism, but only the former have a chance of ruling out Gettier conditions. Second, the other major, broad strategy for dealing with Gettier cases has been to adapt the truth condition of knowledge, either by requiring that the belief not be grounded in any false beliefs or by requiring that there is no true proposition such that adding it to the subject's evidence would preclude the subject's belief from being justified [40] Although satisfaction of the truth condition on knowledge is generally not attributable to the subject, this is not any reason to reject attributabilism, as I have argued above [41] If the nonattributability of the truth condition is not a problem, it is easy to infer that there is similarly no problem with adapting the truth condition to rule out Gettier cases [42] As with the externalist strategy above, then, there is no indication that epistemologists attracted to a view of this sort have found any reason to reject attributabilism

---

[38] I am grateful to Jennifer Lackey for this objection.

[39] Goldman, "Discrimination and Perceptual Knowledge", Nozick (*op. cit.*), Plantinga, *Warrant and Proper Function*, Greco, "Knowledge as Credit for True Belief", and Pritchard (*op. cit.*). Although Plantinga acknowledges that "quasi-Gettier cases" may be posed for his view, he says that "Gettier problems afflict internalist epistemologies, and they do so essentially. … The real significance of Gettier problems … is that they show justification, conceived internalistically, to be insufficient for warrant" (*op. cit.*, p. 36).

There are also two points in response to the third objection, which recommends revising attributabilism so that it applies only to justification and not to knowledge First, although it might seem obvious that the excellence of a person's cognitive performance is attributable to her—especially if we hold the broad, world-involving conception of what it means to be a person in the cognitive sense sketched in section II above—this is not always so Suppose Carla is a student from Turkey who has come to study in the United States Prior to leaving Turkey, Carla had avidly read ten different Turkish newspapers and knew that each of them was a reliable source of news Now, however, she can afford only one subscription Just after Carla left Turkey, some very clever propagandists took over nine of the ten newspapers and begin to print only false but plausible stories As luck would have it, Carla happens to choose the one newspaper that still publishes true stories, though she has no more reason to choose it than any of the other papers The beliefs she acquires from the newspaper are true and reliably formed and well-supported by evidence (that is, she has good evidence both for the particular beliefs and for the source from which they derive), but they are not knowledge The problem is not merely that Carla's beliefs are accidentally *true*, but also that they are accidentally *justified* [43] Although her cognitive performance is excellent in both an externalist sense and in an internalist sense, this is a mere accident Instead of having lots of justified, true beliefs, Carla easily could have had lots of unjustified, false beliefs What makes for the difference between these two cases is not something that is attributable to her The upshot, then, is that even if we restrict attributabilism to justification, we will still have problems ruling out accidentality The skeptical argument I have sketched above could therefore be adapted to apply even to views that accept only the weaker form of justification-attributabilism

---

[40] For one example of the first strategy, see Michael Clark, "Knowledge and Grounds A Comment on Mr Gettier's Paper," *Analysis*, XXIV (1963) 46–48, and for an example of the second, see Peter Klein, "A Proposed Definition of Propositional Knowledge," this JOURNAL, LXVIII, 16 (August 19, 1971) 471–82 See Robert Shope, *The Analysis of Knowing* (Princeton University Press, 1983), and "Conditions and Analyses of Knowing," in Paul Moser, ed , *The Oxford Handbook of Epistemology* (New York Oxford, 2002), pp 25–70, for more on the main responses to the Gettier problem

[41] And, I might add, it has obviously not prevented the many epistemologists who explicitly endorse attributabilism (or something even stronger, like responsibilism) from doing so

[42] I should emphasize that I am saying it is an *easy* inference to make—not that the inference is *correct*

[43] For this case and others like it, see my "Accidental Truth and Accidental Justification " I argue there that in some cases even our basic belief sources, like perception, can produce beliefs that are merely accidentally justified

The second response to this objection is that to restrict the scope of attributabilism in this way would ultimately amount to capitulating to the skeptic Michael Williams calls this "the *epistemologist's dilemma*"

> we can either accept skepticism, or make changes in our pre-theoretical thinking about knowledge that shrink the domain, or alter the status, of what we previously thought of as knowledge of objective fact In making such changes, however, we inevitably appear to be making very large concessions to the sceptic Unkindly put, the epistemologist's dilemma is that we can either agree with the sceptic directly, or in a roundabout, grudging way [44]

In rejecting attributabilism—and thereby following the grudging horn of the dilemma—we would in effect be admitting that the sort of relationship with the truth that we currently think knowledge to be is impossible

Knowledge, as we now conceive of it, is something like the nonaccidental possession of truth But the very notion of nonaccidentality presupposes that there *is* a link between what the subject is doing and the goal that is to be accomplished Suppose, for the sake of argument, that there were some nonattributable condition *c* that provided the link between the subject and the truth Whether or not *c* obtained would be a fact that is entirely independent of the subject's justification In that sense, the obtaining of *c* would itself be accidental, relative to the subject's cognitive nature and cognitive performance—nothing in the subject's nature or in what the subject does has any bearing on whether or not *c* obtains Knowledge, understood as depending on some nonattributable condition of this sort, would therefore be accidental as well

To reject or restrict attributabilism, then, is to allow that knowledge amounts to nothing more than the accidental possession of truth That really would be a roundabout way of saying that there is no such thing as knowledge

---

[44] Williams, *Unnatural Doubts* (Princeton University Press, 1996), p 22

V

Traditional (or, as I have called it, "short road") skepticism has typically been grounded in the various skeptical scenarios. Recent discussion of it has largely fallen under one or more of the following three headings. First, traditional skepticism is usually taken to rely on internalist accounts of knowledge and justification, while externalist views are often thought to be immune to traditional skepticism. Second, it is argued that traditional skepticism may be limited by restricting the force of its arguments to just those conversational contexts in which the skeptical scenarios are relevant. And, third, some epistemologists offer a Moorean response to traditional skepticism. As I shall now show, the argument for the *new* skepticism, which results from following the long road, differs significantly from traditional skepticism in every one of these respects.

There are three ways in which the argument for traditional skepticism has been couched in an internalist framework. First, in Barry Stroud's influential presentation of skepticism, he argues that a subject must be able to rule out alternatives incompatible with her belief.[45] Suppose that an airplane spotter in wartime has been trained to recognize airplanes as belonging to type F just in case they have features *x*, *y*, and *z*. The training is very useful in that it allows the spotter to pick out dangerous airplanes with relative ease. But the spotter has not been told about the existence of airplanes of type G, which also have features *x*, *y*, and *z*. Stroud argues that, even though the spotter is doing everything required by her training, "it is nevertheless a condition of knowing that the plane is an F on the basis of *x*, *y*, and *z* that one know it is not a G" (*op. cit.*, p. 69). The point can then be adapted to a general argument for skepticism: in order to know anything about the external world (for example, that there is an owl in the tree), I must rule out the possibility that I am having a dream that has the same experiential character as the purported perception. But I do not know that I am not dreaming, so, I cannot know anything about the external world. Response to this presentation of traditional skepticism has tended to focus on whether Stroud is correct in saying that "Facts about the way we speak and the procedures we follow in everyday life do not show that the sceptical philosopher has misunderstood or distorted the nature of knowledge" (*op. cit.*, p. 69). In *Unnatural Doubts*,

---

[45] Stroud, *The Significance of Philosophical Skepticism* (New York: Oxford, 1984), pp. 67–70. See Bryan Frances, *Skepticism Comes Alive* (New York: Oxford, 2005), for a presentation of a different take on this sort of skepticism, which requires only that one rule out the "live" possibilities incompatible with one's belief—that is, those possibilities that at least some experts believe to be true.

Williams argues at length that Stroud's skeptical argument depends on the acceptance of what he calls *epistemological realism*—that is, the view that there are objective hierarchical relations of evidence, which do not vary from one context to another. He objects, then, that this elaborate epistemological view is an unnecessary and, given the fact that it leads to skepticism, unwelcome theoretical burden. Williams, and some other epistemologists, favor a *contextualist* approach, which takes the standards of evaluation for a subject's belief to vary from one context to another. So, although it might sometimes be true that the airplane spotter needs to rule out the possibility that the plane is a G, this is not always the case. (I shall say more below about the version of contextualism that takes the relevant context of evaluation to be that of the person who attributes knowledge to the speaker.)

The second way in which traditional skepticism has been linked to internalism can be seen in Richard Fumerton's formulation of the basis for skepticism.[46] He argues that the skeptical scenarios, like Descartes's evil demon or the brain-in-a-vat hypothesis, are important, not because they are possibilities one needs to rule out in order to have knowledge, but because they show that our justification for beliefs about the external world (the past, other minds, and so on) is *indirect* (*op. cit.*, pp. 40–41). Given that my experience as if I am sitting at a desk now could be produced in me by an evil demon, I cannot be directly aware that I am sitting at a desk. So, I must be engaging in some kind of inference, connecting my current experience with the belief that I am sitting at a desk. Fumerton then argues that the traditional skeptic relies on what he calls the principle of inferential justification (*op. cit.*, p. 36):

> To be justified in believing one proposition *P* on the basis of another proposition *E*, one must be (1) justified in believing *E* and (2) justified in believing that *E* makes probable *P*.

The skeptic argues that, for almost all of our beliefs, we fail the second clause. There is no deductive connection between our beliefs and the evidence on which they are based, and we are in no position inductively to infer the truth of our beliefs from our evidence. As Fumerton acknowledges, externalists will almost certainly reject at least the second clause of this principle.[47] A reliabilist, for example, will say that an inferred belief *P* will be justified as long as the premise belief *E* is justified and the inferential process that allows the subject to derive *P* from *E* is reliable. There is no need for the subject to also be aware of the reliability of that process, or even aware of the fact that *P* is inferentially derived from *E*.[48] Externalists thus have no trouble with traditional skepticism. Fumerton argues, however, that the very ease with which externalists can refute skepticism shows that their conception of justification is on the wrong track (*op. cit.*, pp. 168–80).[49]

---

[46] Fumerton, *Metaepistemology and Skepticism* (Lanham, MD: Rowman and Littlefield, 1995).

[47] I should also note that direct realists like McDowell will reject the first step of the argument, which seeks to establish that our knowledge of the external world is indirect; see his "Criteria, Defeasibility, and Knowledge," *Proceedings of the British Academy*, LXVIII

A third way in which traditional skepticism is linked with internalism can be found in Pritchard's account of epistemic luck. Pritchard distinguishes between *veritic epistemic luck*, which occurs when "It is a matter of luck that the agent's belief is true," and *reflective epistemic luck*, which occurs when, "Given only what the agent is able to know by reflection alone, it is a matter of luck that her belief is true" (*op. cit.*, pp. 146, 175). Pritchard takes justification in the internalist sense to require reflective access to the grounds for one's belief. This provides an opening for traditional skepticism because there is no way to tell from the inside—that is, on the basis of what is reflectively accessible—whether one is the victim of an evil demon, one cannot have internalistic justification for one's beliefs about the external world. Pritchard concludes, "The skeptical problem is thus, at root, not concerned with the possession of knowledge *simpliciter* at all, but rather with the possession of knowledge of a certain internalist form" (*op. cit.*, p. 220). So, reflective luck is ineliminable; recognition of this fact leads to what Pritchard calls *epistemic angst* (*op. cit.*, p. 225). All is not lost, however, for traditional skepticism does not pose a problem for externalist theories of knowledge and justification.

The immunity of externalism to traditional skepticism has been noted by other philosophers as well. For example, Christopher Hill argues that process reliabilism both provides a correct, nonskeptical account of justification and knowledge and explains why traditional

---

(1982): 455–79. For an argument against the direct realist response to skepticism, see my "The Stoics' Account of the Cognitive Impression," *Oxford Studies in Ancient Philosophy*, xxiii (2002): 147–80.

[48] For two different examples of this sort of view, see Goldman, "What Is Justified Belief?" in G. Pappas, ed., *Justification and Knowledge* (Boston: Reidel, 1979), pp. 1–23, and Alston, "An Internalist Externalism," *Synthese*, lxxiv (1988): 265–83.

[49] For two different arguments that the ease with which externalism can reject traditional skepticism is problematic, see the new evil demon problem in Cohen, "Justification and Truth," *Philosophical Studies*, xlvi (1984): 279–95, and the contextualist attempt to explain the appeal of skepticism in Keith DeRose, "Solving the Skeptical Problem," *Philosophical Review*, civ (1995): 1–52.

skepticism is plausible but wrong.[50] Nozick motivates acceptance of his counterfactual tracking account of knowledge by pointing out the way in which it allows us to retain everyday knowledge of the world around us, even though we cannot know that we are not brains in vats (*op. cit.*). And Sosa has argued that the traditional skeptical problem of the criterion, which shows that any attempt to prove one's own cognitive faculties are reliable will be epistemically circular, can be avoided by the recognition of an externalist dimension of knowledge.[51]

There are, of course, important differences between these externalist responses to traditional skepticism. I am ignoring them here, however, because all that matters for my purposes is what they have in common—namely, the insistence that there are relations linking knowers to what is known other than the merely evidential. Whether the cognitive link between the subject and the object of her knowledge is established through evidence or through reliable processes, counterfactual tracking, or intellectual virtues, all of the views in question fall under the general rubric of fallibilism. For that reason, all of them are affected by the argument for the new skepticism.

It should similarly be clear that the new skeptic does not rely on internalism in any of its forms. That is to say, it does not require acceptance of Stroud's "ruling-out" principle (and, therefore, does not imply any commitment to epistemological realism). It does not rely on Fumerton's principle of inferential justification (or, for that matter, on any other inferential principle, such as the principle of epistemic closure[52]). And it does not require that justification be tied to what is reflectively accessible to the subject. The argument I have sketched does not lead to epistemic *angst* but to skepticism, for what is at stake is not merely what Pritchard calls *reflective* luck, but rather *veritic* luck. No matter how excellent the person's cognitive performance with respect to her true belief that *p* may be, it is still possible that her belief is true by accident. There is no way to rule this out while also requiring knowledge to be attributable to the person who has it. So, every kind of fallibilism—both internalist and externalist—has to confront the problem of accidentality (that is, veritic luck). And every kind of fallibilism—both internalist and externalist—will be unable to solve that problem while remaining committed to attributabilism. Therefore, every kind of fallibilism—both internalist and externalist—will fall prey to the new skepticism.

---

[50] See his "Two Cheers for Process Reliabilism," *Pacific Philosophical Quarterly*, LXXV (1994): 12–28, and his "Process Reliabilism and Cartesian Skepticism," *Philosophy and Phenomenological Research*, LVI (1996): 567–81.

[51] See his "How to Resolve the Pyrrhonian Problematic."

[52] In "Epistemic Operators," this JOURNAL, LXVII, 24 (December 24, 1970): 1007–23, on p. 1011, Fred Dretske says of the closure principle (which he calls the "thesis of penetrability") that "Almost all skeptical objections trade on it." I am not sure that this is true, but in any case it is clear that the new argument for skepticism does not require it.

Here is another way of thinking of the broad dialectic involving traditional skepticism, externalism, and the new skepticism. In essence, a traditional skeptic is arguing that, with respect to what is reflectively accessible to the subject, there is no difference between someone (S_1) who is the victim of an evil demon and someone (S_2) who is not. They have all of the same experiences and can make use of all the same principles of inference. The traditional skeptic concludes, then, that there is no *epistemic* difference between them; either S_1 cannot know that she is not the victim of an evil demon, and, for that reason, cannot know anything about the world in which she lives. Because she is in the same epistemic position as S_1, S_2 also cannot know that she is not herself the victim of an evil demon and cannot know anything about her own environment. To this argument, the externalist replies that there *is* an epistemic difference between the two people: S_2 has intellectual faculties that actually do cognitively fit her with her environment.[53] This is a fact about her that distinguishes her from S_1, even if she herself does not see it.[54] Now, here is where the new skepticism differs importantly from traditional skepticism. Suppose it is granted to the externalist that S_1 and S_2 are epistemically distinguishable—that is, it is granted that S_2 has a cognitive character that supervenes on world-involving properties like the reliability of her faculties. The new skeptic argues that, even when we take all of those externalist features into account, S_2 is indistinguishable from someone (S_3) who has reliable faculties and a true belief but is in a Gettier case (or a case of accidental justification). In other words, S_2 and S_3 are indistinguishable with respect to every feature of their cognitive performance, except for those that are entirely extrinsic to them. But S_3 does not have knowledge, so S_2 cannot have knowledge, either. Because *everything* relevant to her cognitive performance has already been taken into account, there simply is nothing left—no matter how externalist—by which we can distinguish S_2 from S_3. The upshot, then, is that externalism—no matter how well it fares as a response to traditional skepticism—is just as vulnerable to the new skepticism as internalism is.

---

[53] See, for example, Sosa's argument that the aetiology of a belief can make a difference to its justification, in BonJour and Sosa, *Epistemic Justification: Internalism vs. Externalism, Foundations vs. Virtues* (New York: Oxford, 2003), p. 151.

[54] The internalist and the traditional skeptic will reject this move. But, regardless of whether one accepts it, the externalist's conception of what it is to be an epistemic subject is at least philosophically plausible. I have suggested that it is importantly similar to the way in which compatibilists conceive of what it is to be a practical agent; even incompatibilists who reject that conception are nevertheless compelled to recognize the intuitive pull of the view.

So much, then, for internalism and externalism. The second general response to traditional skepticism that has been prominent recently is attributor-contextualism.[55] According to this view, whether a subject may be properly said to have knowledge in a given case is determined in part by contextual features of the person who attributes the knowledge to the subject. Thus, contextualists focus on knowledge attributions rather than on knowledge itself. They argue that the semantic value of a knowledge attribution like, "John knows where his keys are," will change from one context to another, depending on which epistemic standards are operative in the attributor's context. If, for example, the attributor is in a conversation where the skeptical scenarios are relevant to the discussion—as may happen, perhaps, in a philosophy class—the epistemic standards by which John's cognitive performance is measured will be quite stringent. Because John's cognitive performance is quite ordinary, the attributor's claim would be false in that context. If, however, the attributor is in a very different sort of conversation—say, she's leaving a party and needs to get a ride with John—the relevant epistemic standards will be quite relaxed. Her attribution of knowledge to John will then be correct.[56]

Contextualists take one of the strengths of their view to be the delicate balance it strikes between skepticism and anti-skepticism. The intuitive force of skepticism is honored in that the skeptic is actually correct, at least some of the time. When skeptical scenarios and skeptical arguments are relevant in the context of a knowledge attribution, the attribution is almost certainly false. But, on the other hand, the intuitive appeal of common sense is honored, too. Our everyday talk about knowledge is generally right, if we do not think about it too much. Because the vast majority of our time is spent in ordinary contexts, our knowledge attributions are largely correct. For this reason, contextualism can be taken as offering a vindication of anti-skepticism.

---

[55] See, for example, Cohen, "How to Be a Fallibilist", DeRose (*op. cit.*), and David Lewis, "Elusive Knowledge," *Australasian Journal of Philosophy*, LXXIV (1996): 549–67. Attributor-contextualism is to be distinguished from a view that takes the relevant context to be determined by situational or disciplinary aspects of the knowing subject—for example, a doctor would be evaluated in accordance with the standards of the medical community; for this view, see David Annis, "A Contextualist Theory of Epistemic Justification," *American Philosophical Quarterly*, XV (1978): 213–19, and Williams, *Unnatural Doubts*.

[56] I am of course ignoring many of the details of contextualism (for example, an accounting of the mechanisms by which epistemic standards are raised), which have been developed in different ways by the different advocates of the view, as they are not relevant to my purposes here.

What is crucial to contextualism is the idea that epistemic standards can go up and down from context to context. On at least Stroud's presentation of traditional skepticism, we can see the skeptic as arguing that one's cognitive performance is good in some ways, but just not good enough for knowledge. The subject's performance may eliminate some possibilities incompatible with her belief, but others will remain. It is at least plausible that the contextualist strategy will work in response to this type of traditional skepticism.[57] But, it should be clear, the strategy has no bearing on the argument for the new skepticism. We can suppose the subject's cognitive performance to be as good as possible, consistent with fallibilism. There is still a gap between the subject's performance and the truth, which cannot be bridged by anything the subject does. No matter how we raise or lower the relevant epistemic standards, that gap remains.

Contextualism is powerless to reply to the new skepticism because it has nothing to say about how accidentality (that is, either accidental truth or accidental justification) is ruled out. There are some constraints on knowledge attributions that are context-invariant. That the belief in question must be true is one of them. It does not matter what epistemic standards are operative in a particular conversation, any attribution of knowledge to a subject who has a false belief will be false. The same goes for conditions meant to rule out luck or accidentality. As Cohen acknowledges, it is "very implausible that we can truly ascribe knowledge in the Gettier case in *any* context."[58] The person who just happens to look at the one real barn in a county littered with fake barns simply does not bear the right relation to the truth, no matter how good her cognitive performance might otherwise be. The new skeptic argues that even a subject who purportedly has knowledge also fails to bear the right relation to the truth. Because the problem is categorical, and not a matter of degree, the basic contextualist maneuver offers no hope of defense against the new skeptic.

The third general response to traditional skepticism has been the revival of Moorean common sense.[59] Moore's basic strategy is to meet the skeptic's argument obliquely: rather than identifying a false premise or an invalid inference, he says that the premises of the skeptic's highly abstract argument can never be known with as much certainty as we have for our ordinary beliefs about the world around us [60] Thus, responding to Russell's argument that our knowledge of the external world is less than certain, Moore says

---

[57] However, I should note that for Fumerton's version of traditional skepticism, there is no clear sense in which lowering the relevant epistemic standards would make any difference to the skeptic's argument. The problem, as Fumerton presents it, is not that our cognitive performance is not good enough but rather that it just is not good *at all*.

[58] Cohen, "Contextual Solutions to Epistemological Problems: Scepticism, Gettier, and the Lottery," *Australasian Journal of Philosophy*, LXXVI (1998): 289–306, on p. 299; emphasis in the original. Cohen is replying—convincingly, to my mind—to Lewis's attempt (*op. cit.*) to offer a contextualist solution to the Gettier problem.

[59] Moore's argument strategy can be seen in his papers, "Proof of an External World," "Four Forms of Scepticism," and "Certainty," in his *Philosophical Papers* (London

> Russell's view that I do not know for certain that this is a pencil or that you are conscious rests, if I am right, on no less than four distinct assumptions And what I can't help asking myself is this Is it, in fact, as certain that all these four assumptions are true, as that I *do* know that this is a pencil and that you are conscious? I cannot help answering It seems to me *more* certain that I *do* know that this is a pencil and that you are conscious, than that any single one of these four assumptions is true, let alone all four [61]

So, skepticism is ruled out from the very beginning Moore is not really responding to the skeptic—he is telling us to just ignore her Implicit in this strategy is the belief that philosophy is simply incapable of upsetting the convictions of common sense As William Lycan argues, "there is no such thing as a skeptical argument that does not [appeal to some abstract epistemological premise] Which is to say that *there is no good reason to accept the argument*, the unargued principle is only philosophy stuff" (*op cit*, p 41)

George Allen and Unwin, 1959), "Hume's Theory Examined," in his *Some Main Problems of Philosophy* (New York Collier Books, 1953), and "Some Judgments of Perception," in his *Philosophical Studies* (London Routledge, 1922) The revival of the Moorean approach can be seen in, for example, William Lycan, "Moore against the New Skeptics," *Philosophical Studies*, CIII (2001) 35–53, Noah Lemos, *Common Sense A Contemporary Defense* (New York Cambridge, 2004), and James Pryor, "What's Wrong with Moore's Argument?" *Philosophical Issues*, XIV (2004) 349–78 Some epistemologists adapt Moore's argument to an externalist framework, see Sosa's "Reflective Knowledge in the Best Circles," and "How to Defeat Opposition to Moore," *Philosophical Perspectives*, XIII (1999) 141–53, Michael Bergmann, "Epistemic Circularity Malignant and Benign," *Philosophy and Phenomenological Research*, LXIX (2004) 709–27, and Pritchard (*op cit*) Pritchard, however, acknowledges that his neo-Moorean proposal does not solve the problem of reflective epistemic luck and thus is compatible with our being susceptible to epistemic *angst* See my "Epistemic Circularity Squared? Skepticism about Common Sense," for a response to Bergmann's Reidian-inspired defense of common sense See Frances (*op cit*, pp, 164–77), for a compelling skeptical reply to the Moorean approach

[60] According to Frances, the "surprising" result of this oblique strategy is "that we can defeat the sceptic without doing the real work we can defeat her without showing just what it is that makes it true that Joe knows he has hands or even showing where the sceptic's argument has gone wrong" (*op cit*, p 167)

[61] Moore, "Four Forms of Scepticism," p 226, emphasis in the original Moore is responding to Bertrand Russell, *An Outline of Philosophy* (London George Allen and Unwin, 1927), I elide the four assumptions Russell makes as the particular details of his argument are irrelevant both for Moore's purposes and for my own

I am not at all sympathetic to this sort of pessimism about philosophy, in large part because there is no clear line dividing philosophy from common sense This is obvious from the way in which the dialectic between Moore and the new skeptic would continue We can begin by adapting Moore's argument so that he is saying that he is more certain that he is holding a pencil and that he has hands than he is that either fallibilism or attributabilism is true There are then two ways in which this claim might be understood On the one hand, we could take Moore to be saying that he has a philosophical account of what certainty is and that his belief that he is holding a pencil comes out better, given the particulars of that account, than does either fallibilism or attributabilism On the other hand, we could understand Moore to be saying merely that he is more certain, in some intuitive sense, that he is holding a pencil than he is with respect to the two philosophical theses

In response to the first interpretation of Moore's claim, the new skeptic will point out that Moore has not offered any serious characterization of what either certainty or knowledge amount to If he had, we would be able to tell whether Moore was committed to the theses of fallibilism and attributabilism If in fact Moore is committed to the combination of fallibilism and attributabilism, it does not matter whether his purported knowledge that he is holding a pencil meets the epistemic standards specified by his view better than do the theses that constitute his view[62] His philosophical account would still be vulnerable to the new skepticism, and for that reason his claim to certainty about holding a pencil would be undermined

On the other hand, the new skeptic might actually agree with Moore's claim if it is taken in the second way From an intuitive point of view, it is just obvious that we are more certain that we have hands than that some abstract, difficult to comprehend philosophical theses are correct But the problem with claims from the intuitive point of view is that they are not yet *philosophically respectable* There are many commonly accepted ways of speaking, which prove to be quite useful in daily life, but nevertheless cannot be taken at face value when one tries to understand how things really are A classic example is our usual way of referring to the waxing and waning of sunlight as the rising and setting of the sun, despite the fact that the change of light results from the motion of the earth, not the sun In some cases, as with our talk of sunsets, the intuitive way of speaking can be corrected by our broader understanding of the relevant theory In other cases—for example, the way it seems intuitively obvious to bigots of various stripes that people of their own preferred groups are superior—the relevant scientific or philosophical theory shows that the way things look from the intuitive point of view is utterly wrong So, with respect to the intuitive claim that we can be certain that we have hands, which is it? Is this mostly correct, or not? The only way to find out is to attempt to make the claim philosophically respectable But if we do not have a workable epistemology—as the new skeptic argues—the only proper conclusion to draw is that our attributions of knowledge are not part of our best understanding of the world We can make them all we want in daily life, but we cannot take knowledge attributions seriously when we try to say how things really are[63] To the extent, then, that one wants to be a philosopher—rather than merely someone who is comfortable in her own opinions—the new argument for skepticism cannot be avoided

---

[62] If Moore is *not* committed to either fallibilism or attributabilism, he would escape the new argument for skepticism He would, however, have to face the objections to either infallibilism or to the rejection of attributabilism mentioned above, depending on what shape his view took The road to skepticism may be long, but the destination seems to be inevitable

VI

Here is where this stretch of the long road ends Given that just about every epistemological view that is a going concern is committed to both fallibilism and attributabilism, the conclusion is that knowledge, as we now conceive of it, is impossible For the first time, then, we have a skeptical argument that applies equally well to internalism *and* to externalism Unlike the traditional, short road versions of skepticism, it does not in any way presuppose an internalist conception of knowledge or of justification It does not make use of any principles of inference that externalists are likely to reject It does not require the subject to rule out skeptical scenarios incompatible with her beliefs It cannot be answered by arguing that contextual features of the person making knowledge attributions raise and lower our epistemic standards for what may properly be said to be knowledge And it cannot be avoided by retreating to the safety of common sense

The skepticism that results from following the long road is provisional in a way that the short road version is not, depending as it does on the absence of a workable epistemology But this is also a skepticism without remainder it embodies no controversial conceptions of knowledge (except, of course, those it takes from its opponents) And as such, paradoxically, it may be the most reasonable form of skepticism of all

---

[63] Just as a moral error theorist may still hold that it is practically useful to make false moral judgments, so, too, it may be practically useful to make knowledge attributions even if skepticism is correct On this point, see Frances (*op cit*, pp 88–97)

BARON REED

Northwestern University

BOOK REVIEWS

Happy Lives and the Highest Good: An Essay on Aristotle's Nicomachean Ethics. GABRIEL RICHARDSON LEAR. Princeton: Princeton University Press, 2004. 238 p. Cloth $35.00.

Even a reasonably attentive reader of the *Nicomachean Ethics* may be surprised when Aristotle announces, with barely ten pages to go, that happiness (*eudaimonia*) consists exclusively in theoretical contemplation (*theôria*). For the previous couple of hundred pages, he appeared to be arguing that the happy life is one of political engagement, primarily consisting of morally virtuous acts, true friendship, and some theoretical contemplation. It seemed to be the task of a carefully molded cognitive and desiderative capacity for excellent practical judgment (*phronêsis*) to determine, manage, and order these different activities properly so as to constitute a happy life. Suddenly, however, Aristotle says that such a life is happy only "in a secondary way" (1178a9). What had looked like an inclusive, attractive, and so reasonably plausible conception of human happiness appears to be rejected in favor of a monistic, intellectualist, and implausible one.

Gabriel Richardson Lear addresses this notorious puzzle about the *NE* and offers a clever solution to it in her original and well-argued book. A significant bonus of her treatment is that it connects Aristotle's ethical views with his physics and metaphysics more generally and with their Platonic antecedents in illuminating and insightful ways, which will surely spur further research. She defends the most straightforward reading of the text: Aristotle's conception of *eudaimonia is* monistic and his exclusive identification of it with theoretical contemplation in X.7 is what he has intended all along. *Theôria*, as the highest good, confers value on subordinate goods, particularly including moral virtue, whether or not we know it does.

As Lear is well aware, many scholars reject a monistic interpretation because of the problem that the identification of *eudaimonia* and contemplation generates for morally virtuous action. Aristotle believes that morally virtuous action must be chosen for its own sake (1105a31–2). But he also maintains that we choose virtuous actions not only for their own sakes but also for the sake of *eudaimonia* (1097a34–1097b5). If we identify *eudaimonia* with the monistic end of *theôria*, then we must choose morally virtuous action for the sake of *theôria*. This appears problematic for two reasons: (1) it seems simply

0022-362X/07/0405/263–68

false that morally virtuous action is a means to *theôria*, indeed, it is more plausible to think that engaging in morally virtuous actions would be a *hindrance* to *theôria*, and (2) if we choose morally virtuous actions for the sake of *theôria*, it would seem to conflict with the claim that the truly virtuous person engages in virtuous actions *for their own sake* Many readers believe that the best resolution of these problems is to interpret *eudaimonia inclusively* as the set of all intrinsically valuable goods Morally virtuous activity and *theôria*, as intrinsically valuable, would each be components of *eudaimonia* Therefore when Aristotle says that we engage in morally virtuous action "for the sake of" *eudaimonia*, he is not referring to a means/end causal relationship, but to a constituent/whole relationship

The opening chapters of Lear's book are critical and explain why an inclusivist reading is ultimately inadequate to the requirement of being a truly "most final" good, which requires that a "more final" end play a value- and form-conferring role on subordinate ends (18, 42) Lear argues persuasively that if *eudaimonia* consists in a set of intrinsically valuable members, then it is not *eudaimonia itself* that determines *what* we ought to pursue nor does *it* explain why subordinate goods have value, and so eudaimonia fails to play its necessary form- and value-conferring roles

To defend a monistic intellectualist conception of *eudaimonia* Lear must find a way of understanding what she calls "middle-level ends," such as morally virtuous action, which are desirable both for their own sakes and for the sake of *eudaimonia* Her solution is to argue that Aristotle has a third "for the sake of" relation besides means/end and constituent/whole what she calls "teleological approximation" (c 4) Lear argues that when Aristotle says that mortal creatures act "for the sake of" the divine by reproducing, he is best understood as saying that mortal creatures "imitate" or "approximate" the divine by reproducing Approximation is also central to acting "for an object of love" as described in Plato's *Symposium* (the topic of an Appendix) when a person acts for love, he imitates and approximates the object of his love Similarly, on Aristotle's world-view, the natural world seeks to imitate the divine Prime Mover, who is pure contemplative activity Carried over to the *NE*, human beings imperfectly attempt to imitate and approximate, as far as they are able, the activity of divine contemplation of the Prime Mover by engaging in contemplation themselves Lear's ingenious addition to this account is that just as the capacity for human contemplation approximates divine contemplation, so morally virtuous activity (via the activity of *phronêsis*) approximates theoretical contemplation In a means/end relationship the end sets the normative standard for what is going to count as successful means and, furthermore, the value of the means is conferred by the value of the end—the value of horseback riding confers value on the activity of bridle making. The same relationships are operative in teleological approximation: *y* confers value on and determines the conditions of success for *x*, which is done "for the sake of *y*." In order for Lear to be able to resolve the problem of "middle-level ends" the *intrinsic* value of theoretical contemplation must be passed on to the approximation of contemplation that is morally virtuous action. This is the most central and controversial part of the book.

Lear attempts to establish two things: (1) when *x* teleologically approximates *y* and *y* is intrinsically valuable, the intrinsic value of *y* is inherited by *x* and *y* is the source of *x*'s value (chapter four, esp. 83–88), and (2) morally virtuous action as an exercise of *phronêsis* is indeed a teleological approximation of theoretical contemplation (chapter five). I found (2) more persuasive than (1). In brief, *phronêsis* approximates *theôria* insofar as *phronêsis* seeks truth in the practical sphere of action, while *theôria* seeks truth about theoretical matters (103). Although Lear admits that Aristotle never explicitly *says* that *phronêsis* gets its value from its kinship with *theôria* (90), she makes a fairly persuasive case that there is such a relationship. (1), however, is more problematic, yet just as essential to Lear's argument. Not only must the paradigmatic end confer *value* on the approximating thing, it must also confer *intrinsic* value if teleological approximation is to solve the problem of how morally virtuous activity could be chosen for its own sake and for the sake of *theôria*. Why should something's being like something else give it intrinsic value, particularly if, as Lear insists, it is simply the value of the most final end that confers value on the subordinate, approximating ends and *not* some quality of it (83–84)? There are complex issues here concerning the difference it might make whether the object of approximation is an *object*, as is suggested by comparison with Plato's Forms, or an *activity*, as is more in line with Aristotle's idea of *theôria*, but Lear herself speaks of objects and their "characteristics" (83–84). Further, the more one insists that the value of the approximating object is *intrinsic* to it, the less it would seem to depend on a relationship of teleological approximation or of any other sort to some other object. Correspondingly, to the extent that the approximating object's value does depend upon its relationship to another object, it is difficult to understand how its value could be intrinsic and not extrinsic.

Chapter six concerns the nature of moral virtue and "the fine (*to kalon*)" and chapter seven examines three virtues in detail—courage, temperance, and greatness of soul. I found these chapters, although independently interesting (especially the discussion of the fine), a bit tangential to the overall thesis of the book. In particular they do not help to show how the relationship between morally virtuous action and *theôria* is one of teleological approximation. I had been expecting that in the detailed consideration of particular virtues we would see textual evidence that Aristotle thinks that acting in these ways (for instance, courageously) approximates the activity of *theôria*. But in fact that claim, as far as I can tell, rests *entirely* on the argument of chapter five, where Lear attempts to show that *phronêsis* is a teleological approximation of *theôria*. Insofar as courageous or temperate activities are displays of *practical wisdom, to that extent* they are approximations of *theôria*. In the specific discussions of the virtues, however, all that Lear shows is that we can make sense of the value of those virtuous actions in terms of their contributions to leisure and the fact that they make clear and visible (part of what it means for these actions to be "fine" according to Lear) the agent's commitment to an excellent use of leisure, which is going to be, whether the agent knows it or not, the use of it for *theôria*. So, for example, courageous action in battle is *kalon* because it is a visible example of an agent's commitment to protecting his polis and the leisure it makes possible during peacetime. This is plausible in its own right, but plainly *not* an example of courageous action's being for the sake of contemplation in the sense of being a teleological approximation of *theôria* (except insofar as it is an activity of *phronêsis*), the relationship between courageous action and contemplation in fighting for leisure is means/end, not approximation.

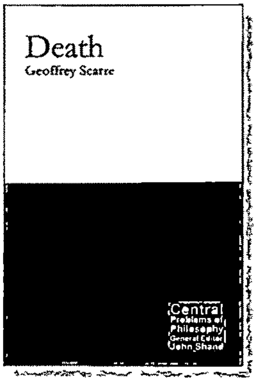
Has Lear solved the monistic reading's problem of the intrinsic value of morally virtuous actions in a philosophically, ethically, and politically satisfying way? Her final chapter is cautious and a bit concessive. If Lear's interpretation is correct, it does explain why a person who performs morally virtuous actions in lieu of the opportunity to contemplate is nevertheless doing something of intrinsic value (since morally virtuous action is "contemplation of a sort" (*theôria tis*, cf. *NE* 1178b32 and Lear, 91–92, 194–96)), but it does not show that an agent should ever *choose* morally virtuous action *over* contemplation. Lear concludes on her final page: "So nothing *prevents* the philosopher from valuing morally virtuous action" (207, my emphasis), which echoes her initial claim that once her argument is complete the *NE*'s conclusion that "the most perfect happiness is the monistic good of contemplation will not strike us as *so problematic*" (4, my emphasis).

Let us grant that Lear has given us a neat way of understanding how practically virtuous actions can provide us with happiness "in a secondary way." She has thus shown us why a life of *theôria*, with morally virtuous actions performed whenever one is not contemplating, is *better* than a life of acting contrary to moral virtue when not contemplating. She adds that “even in the best of circumstances, the time we spend engaging our theoretical reason will necessarily be extremely short in comparison with the time we must spend in using our practical reason” (177, cf. 205). But is this true, particularly if I were to be convinced by Aristotle’s account? Certainly in the best of circumstances some of my ancestors have set me up in a stable community, with adequate servants, and I have a permanent post at a research facility. In such a situation I engage in theoretical contemplation (including working together with truly philosophical friends) for as many hours as I can each day until I need relaxation or sleep. Lear works hard to argue that contemplation’s being our most final end need not imply that we always maximize contemplation (201–05), but she nevertheless concedes that “No doubt the philosopher would like to contemplate as much as he can” (204). I am sympathetic with her claim that Aristotle is not going to offer us an “algorithm” (205) for when we should contemplate and when we should engage in morally virtuous actions. But this still does not address the heart of the problem. While securely pursuing our research and discussing our work over “moderate” dinners, many of us feel (or believe that we ought to feel) conflicted about whether we are neglecting a moral duty to participate in reducing the injustices that beset our society and world. It is clear on Lear’s account why, if my city is being overthrown and my research institute about to be turned into a shopping mall, I have (instrumental) reasons for bravely fighting for my polis. As I spend time by thus showing in action my commitment to the value of the leisure necessary for theoretical contemplation, I can console myself that I am, “secondarily,” engaged in “contemplation of a sort.” Nevertheless it still remains, at the end of the day, a consolation prize. Since Lear does not dispute that happiness in the *NE* concerns in the end the happiness of the individual (198 note 47), I would and should want to get back to engaging in the highest good, contemplation, as soon as I can. While I will not begrudge myself the need of sleep for I recognize that I am human, I do not see any active reason why I ought to spend time on the upbringing of my children (especially if I have care providers to do it for me) let alone on trying to rectify the injustices of my society.

Lear says that “although the philosopher longs to use his theoretical reason, it is not possible for him to ignore his practical life” because he “just finds himself” a citizen or a parent, which are natural parts of being human (200). She continues, “. . . to bring things closer to the philosophical life as we know it, we find ourselves members of university departments and must take a stand . . . on hiring decisions, the awarding of grades, and where to take visitors for dinner Since these decisions cannot be avoided, the philosopher *must* decide how to make them" (201) But why can't they be avoided? Aren't they only there because we *are not* convinced by the claim that our individual happiness consists most of all in the monistic end of contemplation? What makes the whole puzzle more than just an exegetical exercise in how to make what Aristotle says consistent struck me as missing from this study This is somewhat more surprising because of Lear's sensitive and important attention to Plato's influence on Aristotle But in Callicles's criticism of Socrates's political ineffectualness in the *Gorgias* and in the *Republic*'s tension about why a philosopher, once in contemplation of the Forms, would ever return to the "cave" and engage in political affairs, we see a serious concern with an issue that is not resolved by discovering that the value of political virtue approximates the best value of *theôria* If one insists on a monistic reading of *eudaimonia*, then one, and only one, end can be the good If *eudaimonia is* theoretical activity, then even if that does not imply that one maximizes contemplation moment by moment, that fact will make a big difference in shaping one's life around contemplation if at all possible, and not around its inferior approximation, moral virtue Being politically engaged is difficult, frustrating, and wildly detrimental to one's time and ability to do theoretical work As is clear to many readers of this journal, one can get a lot of "theoretical work" done in a comfortable academic position, which affords plenty of leisure, while simply ignoring the political and social injustices that are literally (think of many prestigious urban universities in America) occurring right at one's doorstep Were a person to devote herself to those problems, she might well never do philosophy again If one is safe and secure in a place that affords leisure to engage in *theôria*, then, regardless of the injustices that may exist to maintain it or that may be occurring around it, according to Lear's Aristotle one would have a less happy and less valuable life by leaving it

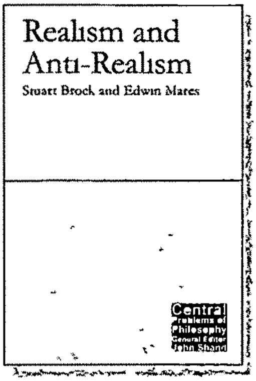
It is a significant strength of Lear's book that even if one is not persuaded that she has resolved the tension between Book X 7–8 and the rest of the *NE*, there is so much of importance that she establishes en route The elaboration of the concept of teleological approximation and the argument that *phronêsis* approximates *theôria* insofar as both seek truth are substantial contributions to our understanding of Aristotle and the unity of his philosophical thinking No one interested in Aristotle's ethics and its connections to his concept of philosophical activity should neglect this book

IAKOVOS VASILIOU
Brooklyn College, The City University of New York



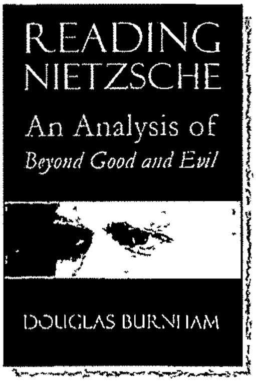
Death
Geoffrey Scarre
Central Problems of Philosophy
978 0-7735-3241-0 $22 95 paper | 978-0-7735-3240-3 $80 00 cloth

"Lucid, informed, and engaging Scarre surveys the literature on the nature and significance of death and makes original contributions of his own "
– Steven Luper, Trinity University, San Antonio


Realism and Anti-Realism
Stuart Brock and Edwin Mares
Central Problems of Philosophy General Editor John Shand

Realism and Anti-Realism
Stuart Brock and Edwin Mares
Central Problems of Philosophy
978 0 7735-3239-7 $22 95 paper | 978-0-7735-3238-0 $80 00 cloth

"Essential reading for anyone who is interested in the contemporary realism/anti-realism debate "
– Eugenio Lombardo, University of Leeds



Reading Nietzsche
An Analysis of Beyond Good and Evil
Douglas Burnham
978-0-7735-3250-2 $22 95 paper | 978-0-7735-3249-6 $80 00 cloth

"An engaging, detailed, and lucid guide to one of Nietzsche's most important works A great introduction to Nietzsche and a sophisticated commentary on *Beyond Good and Evil* "
– Claire Colebrook, University of Edinburgh

MQ UP
McGILL-QUEEN'S UNIVERSITY PRESS
www mqup ca

THE JOURNAL OF PHILOSOPHY

BACK ISSUE SALE

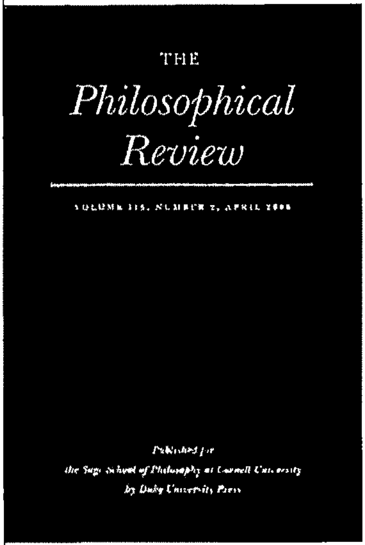
Complete volumes
are now available
at the following discounted prices
in order to reduce our inventory:

Volumes	1 through 98	$10 each
Volumes	99 through 101	$25 each

For more information, contact Pamela Ward,
Business Manager, at pcw1@columbia.edu

Visa, Mastercard, American Express and
PayPal are accepted through our website at
www.journalofphilosophy.org

Duke University Press
is pleased to be the new publisher of
the *Philosophical Review.*


THE
Philosophical Review

*Edited by the faculty of
the Sage School of Philosophy
at Cornell University*

First published in 1892, the *Philosophical Review* publishes original scholarly work in all areas of analytic philosophy, with an emphasis on material of general interest to academic philosophers, and is one of the few journals in the discipline to publish book reviews.

Free online access to all issues back to 2000 is included as part of an individual or student print subscription at **philreview.dukejournals.org**

Subscription Information
Quarterly
Individuals $33
Students $20 (photocopy of student ID required)
Additional postage fees apply for international subscribers

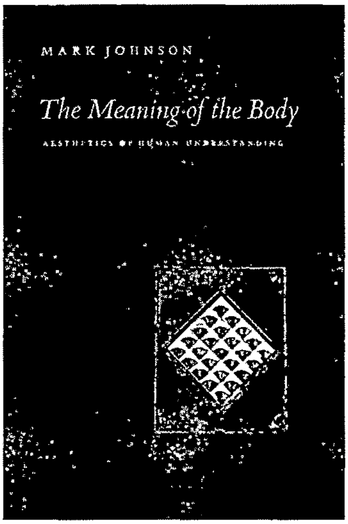
To place your order, please call toll-free 888-651-0122 (in the U.S. and Canada) or 919-688-5134, or e-mail **subscriptions@dukeupress.edu**

For more information, please visit **philreview.dukejournals.org**

DUKE
UNIVERSITY PRESS

PHILOSOPHY

from Chicago



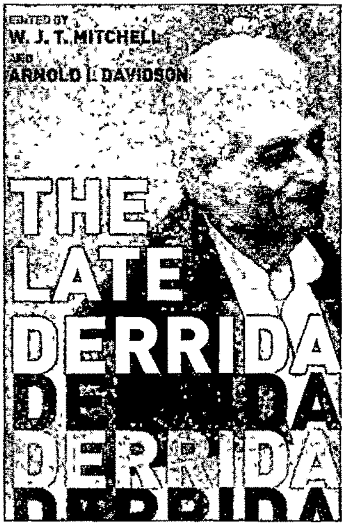
THE MEANING OF THE BODY

Aesthetics of Human Understanding

Mark Johnson

"Grounding his argument firmly in the philosophy of John Dewey, Johnson creates a new vision of the aesthetics of human understanding that is supported by contemporary research from linguistics, psychology, and neuroscience on the embodied nature of human cognition Yet Johnson also provides beautiful examples of how artistic practices exhibit and extend the embodied mind "—Raymond W Gibbs, Jr , author of *Embodiment and Cognitive Science*

CLOTH $32 00



THE LATE DERRIDA

Edited by W J T Mitchell and Arnold I Davidson

The rubric "The Late Derrida," with all puns and ambiguities cheerfully intended, points to the late work of Jacques Derrida, the vast outpouring of new writing by and about him in the period roughly from 1994 to 2004 In this period Derrida published more than he had produced during his entire career up to that point At the same time, this volume deconstructs the whole question of lateness and the usefulness of periodization

UNIVERSITY OF CHICAGO PRESS JOURNALS

PAPER $25 00

THE UNIVERSITY OF CHICAGO PRESS • www press uchicago edu

New from the Center
for the Study of Language
and Information

CSLI
PUBLICATIONS

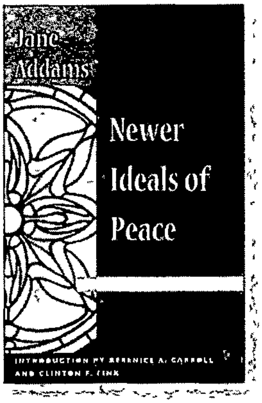
SELF-REFERENCE
Thomas Bolander, Vincent F Hendricks,
and Stig Andur Pedersen
An anthology of previously unpublished essays, *Self-Reference* reexamines the latest theories of self-reference, including those that attempt to explain and resolve the semantic and set-theoretic paradoxes With a thorough introduction that contextualizes the subject for students, this book will be important reading for anyone interested in the general area of self-reference and philosophy
PAPER $27 50

EUCLID AND HIS TWENTIETH CENTURY RIVALS
Diagrams in the Logic of Euclidean Geometry
Nathaniel Miller
Twentieth-century developments in logic and mathematics have led many people to view Euclid's proofs as inherently informal, especially due to the use of diagrams in proofs In *Euclid and His Twentieth-Century Rivals*, Nathaniel Miller discusses the history of diagrams in Euclidean Geometry, develops a formal system for working with them, and concludes that they can indeed be used rigorously
PAPER $27 50

Back in Print
PRACTICAL REFLECTION
J David Velleman
"What do you see when you look at your face in the mirror?" asks J David Velleman in introducing his philosophical theory of action He takes this simple act of self-scrutiny as a model for the reflective reasoning of rational agents our efforts to understand our existence and conduct are aided by our efforts to make it intelligible Reflective reasoning, Velleman argues, constitutes practical reasoning
PAPER $32 50

DISTRIBUTED BY THE UNIVERSITY OF CHICAGO PRESS • www press uchicago edu

UNIVERSITY OF ILLINOIS PRESS



Newer Ideals of Peace

JANE ADDAMS

Introduction by Berenice A. Carroll and Clinton F. Fink

Addams's second book is an unprecedented, revolutionary critique of the pervasive militarization of society. Also included is an extensive, in-depth introduction that provides historical context, analysis, and a reassessment of the theoretical and practical significance of the book.

Cloth, $35.00, Paper, $15.00

Treatise on Love of God

MIGUEL DE UNAMUNO

Translated from the Spanish and annotated by Nelson R. Orringer

This study of religious inwardness proposes to analyze how God can be found within as a beloved person, and serves as prelude to Unamuno's *Tragic Sense of Life*.

Hispanisms

Cloth, $40.00

The Moral Menagerie

Philosophy and Animal Rights

MARC R. FELLENZ

Making use of the critical perspectives found in environmentalism, feminism, and postmodernism, Fellenz argues that taking animals seriously requires a more radical reassessment of our moral framework than the concept of "animal rights" implies.

Illus. Cloth, $60.00, Paper, $25.00

Putting a Face on Things

Studies in Imaginary Materials

MICHAEL CARTER

Taking its inspiration from Gaston Bachelard's idea of a material imagination, this book looks beyond the labor that endures physical survival to the elaboration of the world into forms that accomodate our collective desires.

Distributed for Power Publications

Illus. Paper, $20.00

New in Paperback

Defending Animal Rights

TOM REGAN

Paper, $20.00

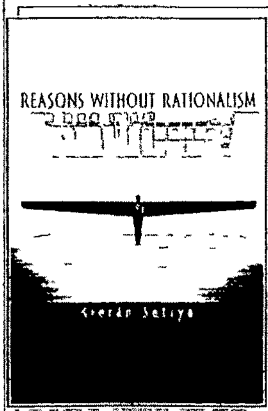
Sport, Play, and Ethical Reflection

RANDOLPH FEEZELL

A *Choice* Outstanding Academic Title

Paper, $20.00

www.press.uillinois.edu • 800-621-2736

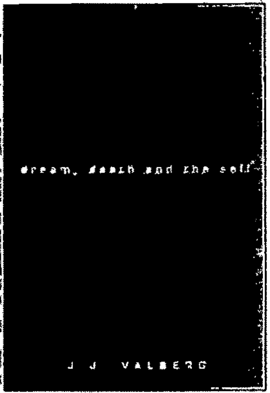


Reasons without Rationalism

KIERAN SETIYA

"Kieran Setiya urges the radical thesis that the 'should' of practical rationality just is the 'should' of ethical character. His argument turns on taking issue with the orthodox view that practical rationality is exercised under the aspect of the good, and on rejecting all forms of the idea that the nature of practical rationality can be understood from reflection about reason and action as such. [H]is book is a pleasure, as well as a challenge, to read."
—John McDowell, University of Pittsburgh

Cloth $29.95 978-0-691-12749-1



Dream, Death, and the Self

J. J. VALBERG

"In Jerome Valberg's extraordinary book, one finds a distinctive conception of philosophical problems, a highly original response to dream skepticism, a deep interpretation of the meaning of death, and a groundbreaking discussion of personal identity. Valberg's position is both striking and masterfully developed."
—Douglas G. Winblad, Vassar College

Paper $29.95 978-0-691-12859-7
Cloth $75.00 978-0-691-12858-0

A New York Times *Editors' Choice*

The Ethics of Identity

KWAME ANTHONY APPIAH

"Suave and discerning."—Jonathan Freedman, *New York Times Book Review*

New in paper $19.95 978-0-691-13028-6

Reference and Description

The Case against Two-Dimensionalism

SCOTT SOAMES

"A text of this scope and care is a must for any program supporting philosophy of language and mind."—*Choice*

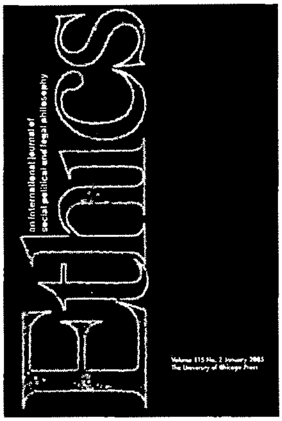
New in paper $19.95 978-0-691-13099-6

PRINCETON
University Press

800-777-4726
Read excerpts online
press.princeton.edu

Ethics

Editor: John Deigh



About the Journal

Founded in 1890, *Ethics* publishes scholarly work in moral, political, and legal philosophy from a variety of intellectual perspectives, including social and political theory, law, and economics. In addition to major articles, *Ethics* also publishes review essays, discussion articles, book reviews, and book notes.

Quarterly
ISSN 0014-1704

For more information, visit
www.journals.uchicago.edu/ET

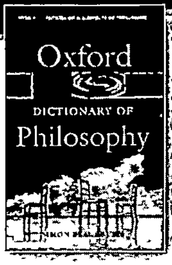
Subscribe today!

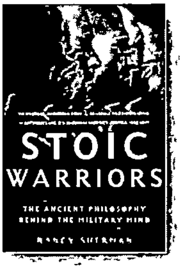
Individual subscription prices:
Rates valid through December 31, 2007. Additional shipping and taxes applied to international orders.

Print and electronic, $48
Electronic-only, $42
Print-only, $42
Special APA, $38
Students (electronic-only), $24
(photocopy of current student ID required)

To Subscribe:
Phone (877) 705-1878 (USA/Canada)
(773) 753-3347 (International)
Fax (877) 705-1879 (USA/Canada)
(773) 753-0811 (International)
Online www.journals.uchicago.edu

PUBLISHED BY
THE UNIVERSITY
OF CHICAGO PRESS
JOURNALS DIVISION
WWW.JOURNALS.UCHICAGO.EDU





New from OXFORD

New in paperback!

The Oxford Dictionary of Philosophy

Second Edition

SIMON BLACKBURN

"Thanks to clever typography, this revision of the 1994 edition eliminates virtually nothing, adding hundreds of new entries in the same 400 pages with no loss of legibility. This edition will double as a world-religion reference, but its original material is, reassuringly, little changed."—*School Library Journal*

(Oxford Paperback Reference)
2006 (paper 2007) 416 pp.
paper $16.95 cloth $45.00

A Future for Presentism

CRAIG BOURNE

How can we talk meaningfully about the past if it does not exist to be talked about? What gives time its direction? Is time travel possible? This defense of presentism—the view that only the present exists—makes an original contribution to a fast growing and exciting debate.

2007 240 pp. $55.00

New in paperback!

Sentimental Rules

On the Natural Foundations of Moral Judgment

SHAUN NICHOLS

"*Sentimental Rules* is hugely original, uncommonly lucid, highly educational, and refreshingly unabashed in its interdisciplinarity. It sets a new standard for naturalized ethics."—Jesse Prinz, University of North Carolina, Chapel Hill

2004 (paper 2007) 240 pp. paper $24.95

A Metaphysics for the Mob

The Philosophy of George Berkeley

JOHN RUSSELL ROBERTS

"The book will, I believe, become a central work in Berkeley scholarship and in the area of early modern philosophy. It is a pleasure to read. It is written in a clear, non-stuffy, often elegant, and sometimes even witty style. It shows an extraordinary grasp of both the primary and secondary literature of early modern philosophy."

—Robert J. Fogelin, Dartmouth College

2007 200 pp. $55.00

Stoic Warriors

The Ancient Philosophy behind the Military Mind

NANCY SHERMAN

"Sherman traces the origins of what we consider soldierlike behavior, reaching back to ancient philosophers like Aristotle, Cicero and Seneca and finding echoes of their moral view in later writers, like Emerson, Stephen Crane and even Adm. James Stockdale, whose account of his years in a Vietnam prisoner-of-war camp cites Epictetus as an ethical guide. The essence of soldierliness is a Stoic ideal, as Ms. Sherman explains: discipline, endurance, a can-do spirit, a stiff upper lip. But she is at pains to show, in wonderfully clear prose, that Stoicism is filled with subtleties and nuance."

— *Wall Street Journal*

2007 256 pp., 15 halftones
paper $15.95

OXFORD
UNIVERSITY PRESS

Prices are subject to change and apply only in the US.
To order, please call 1-800-451-7556. In Canada, call 1-800-387-8020.
Visit our web site at www.oup.com/us/social work

ANNOUNCING:

THE JOURNAL OF PHILOSOPHY

ONLINE

The Journal of Philosophy has added
a print and online subscription option for institutions.

SUBSCRIPTION OPTIONS

Libraries and Institutions (print and online)	$180.00
Libraries and Institutions (print only)	$100.00
Individuals	$45.00
Students, retired/unemployed philosophers	$20.00
Postage outside the U.S.	$15.00

Please consult our website at www.journalofphilosophy.org
for more details

CENTRAL LIBRARY
18.7.08
UNIVERSITY
OF
CALCUTTA
ALIPORE CAMPUS

THE JOURNAL OF PHILOSOPHY

VOLUME CIV, NUMBER 6
JUNE 2007

page
269 *Cartesian Knowledge and Confirmation* Joel Pust
290 *The False Dichotomy between Coherentism and Foundationalism* Sven Ove Hansson
301 *Love and Death* Dan Moller
317 NEW BOOKS ANTHOLOGIES

Published by **The Journal of Philosophy, Inc.**

THE JOURNAL OF PHILOSOPHY

FOUNDED BY FREDERICK J E WOODBRIDGE AND WENDELL T BUSH

Purpose To publish philosophical articles of current interest and encourage the interchange of ideas, especially the exploration of the borderline between philosophy and other disciplines

Editors Bernard Berofsky, Akeel Bilgrami, John Collins, Arthur C Danto, Kent Greenawalt, Patricia Kitcher, Philip Kitcher, Isaac Levi, Wolfgang Mann, Mary Mothersill, Frederick Neuhouser, Christopher Peacocke, Carol Rovane, Achille C Varzi *Consulting Editors* David Albert, James T Higginbotham, Robert May, Charles D Parsons, Wilfried Sieg *Managing Editor* John Smylie

THE JOURNAL OF PHILOSOPHY is owned and published by the Journal of Philosophy, Inc *President*, Arthur C Danto, *Vice President*, Akeel Bilgrami, *Secretary*, Daniel Shapiro, *Treasurer*, Barbara Gimbel, *Other Trustees* Lee Bollinger, Leigh S Cauman, Kent Greenawalt, Michael J Mooney, Lynn Nesbit

All communications to the Editors and Trustees and all manuscripts may be sent to John Smylie, Managing Editor, Mail Code 4972, 1150 Amsterdam Avenue, Columbia University, New York, New York 10027 FAX (212) 932-3721

You may also visit our website at www journalofphilosophy org

THE JOURNAL OF PHILOSOPHY

2007

SUBSCRIPTIONS (12 issues)

Individuals	$45 00
Libraries and Institutions	$100 00
Libraries and Institutions (print and online)	$180 00
Students, retired/unemployed philosophers	$20 00
Postage outside the U S	$15 00

Payments only in U S currency on a U S bank All back volumes and separate issues available back to 1904 Please inquire for advertising rates, ad space is limited, so ad reservations are required

Published monthly as of January 1977, typeset by Dartmouth Journal Services, Waterbury, VT, and printed by The Sheridan Press, Hanover, PA

All communication about subscriptions and advertisements may be sent to Pamela Ward, Business Manager, Mail Code 4972, 1150 Amsterdam Avenue, Columbia University, New York, NY 10027 (212) 866-1742

All materials published in *The Journal of Philosophy* are copyrighted by the *Journal* and are protected by U S copyright law All rights reserved For more information about our copyright policies, please consult our January issue or contact us

The Journal of Philosophy (ISSN 0022-362X) is published monthly by The Journal of Philosophy, Inc , MC4972, Columbia University, 1150 Amsterdam Avenue, New York, NY 10027 Periodicals postage paid at New York, NY, and other mailing offices

POSTMASTER Please send address changes to *The Journal of Philosophy* at MC4972, Columbia University, 1150 Amsterdam Avenue, New York, NY 10027

© Copyright 2007 by the Journal of Philosophy, Inc

ISSN 0022–362X

THE JOURNAL OF PHILOSOPHY
VOLUME CIV, NO. 6, JUNE 2007

CARTESIAN KNOWLEDGE AND CONFIRMATION*

Bayesian approaches to probabilistic inference have long been thought by many philosophers to capture the structure of confirmation in empirical science. Recently, however, such approaches to distinctively philosophical topics have been gaining in popularity. Bayesian conceptions of evidence have been invoked in recent arguments regarding the existence of God, the hypothesis of multiple physical universes, and the Doomsday Argument. Philosophers writing on these topics often claim that, given a Bayesian account of evidence, our existence or something entailed by our existence (perhaps in conjunction with some background knowledge or assumption) may serve as evidence for each of us.

In this paper, I argue that this widespread view is mistaken. The mere fact of one's existence *qua* conscious creature cannot serve as evidence on the standard Bayesian conception of evidence because knowledge of one's existence is a necessary part of the background knowledge relative to which all epistemic probabilities are defined. It follows that any proposition *a priori* entailed by one's existence cannot serve as evidence and that any proposition *a priori* entailed by the conjunction of one's existence and one's other background knowledge cannot serve as evidence, relative to that background knowledge. If I am correct, then many of the aforementioned recent applications of the Bayesian conception of evidence to philosophical problems are undermined. Furthermore, as will become apparent, if I am correct,

---

* I owe a most significant debt to Kai Draper for very many helpful conversations and suggestions about the central ideas of this paper and for saving me from a number of errors. Mark Kaplan and William Wainwright provided extremely useful comments on versions of this paper presented at the 2004 Eastern Division American Philosophical Association Meeting and the 2007 Society for Philosophy of Religion Meeting. I am also indebted to Michael Bergmann, Michael Rea, Paul Draper, Douglas Stalker, Holly Feldheim, David Silver, Christopher Boorse, Jeff Jordan, and Neil Manson for helpful comments.

0022-362X/07/0406/269–89

then the very common assumption that it is possible for every single contingent proposition to serve as evidence, is mistaken

I CARTESIAN KNOWLEDGE AND THE STRONG PROBLEM OF OLD EVIDENCE

The sort of "Cartesian" knowledge which I contend that standard probabilistic approaches to evidential reasoning are unable to treat as evidence includes the following

E_1 = I exist

E_2 = I am conscious

Though I will claim that such propositions are (or should be) certain for a given rational mental subject, I will make no particular assumptions about the ultimate nature of rational mental subjects [1]

Most Bayesians who provide an account of qualitative (incremental or relative) confirmation advance the following "positive relevance" account [2]

PR E confirms H, given background knowledge K, iff $P(H/E\&K) > P(H/K)$

This account of confirming evidence is an attempt to provide a precise characterization of the intuitive idea that E is evidence for H when E makes it *rational to be more confident* that H is true [3] This account also respects the fact that whether a given proposition counts as evidence for a hypothesis depends on one's epistemic situation For example, whether the presence of Ted's fingerprints on the safe is evidence that he is the thief who last night burgled the store depends on whether one knows that Ted is the night manager with frequent and legitimate occasion to open the safe or, instead, a disgruntled former employee with a number of prior burglary convictions PR captures the relevance of one's epistemic situation by relativizing confirmation to background knowledge, K Though I am inclined to think that one's epistemic situation, in the sense relevant to confirmation, contains sensory and other cognitive states which may not be explicitly captured by one's beliefs or knowledge, for present purposes this fact can be accommodated by the idealizing assumption that a fully rational creature is introspectively omniscient, that is, has propositional knowledge of all of its own introspectively accessible mental states

---

[1] Those who claim that such "*de se*" knowledge requires a notion of belief which is fundamentally nonpropositional (see David Lewis, "Attitudes *De Dicto* and *De Se*," *Philosophical Review*, LXXXVIII (1979) 513–43) should be able to recast my arguments in their preferred terms

[2] There is also a Bayesian project of articulating a *quantitative* measure of incremental confirmation If the arguments I advance in this paper are correct, none of the proposed quantitative measures requiring a difference between $P(H/E)$ and $P(H)$, or undefined when $P(E) = 1$, will allow E_1 or E_2 to serve as evidence

[3] Though PR attempts to capture a core intuitive notion of evidence, it also has, as noted by Rudolf Carnap in *The Logical Foundations of Probability*, 2nd edition (Chicago University Press, 1962), pp 382–86, a number of somewhat counterintuitive consequences

On PR, it is clear that K cannot entail E if E is to confirm H because in such circumstances the conditional probabilities in PR will be equal rather than unequal This fact is the basis of the celebrated "Problem of Old Evidence," the first formulation of which is credited to Clark Glymour [4] Given PR and a natural understanding of what an agent's background knowledge or information is, it would seem that if an agent knows E to be the case, then E must have an epistemic probability of 1 and it is impossible for E to confirm (or disconfirm) H since K includes E and hence $P(H/K)$ is identical to $P(H/E\&K)$ The problem is that PR appears to imply the absurd conclusion that once a proposition is known to be true, it can no longer be evidence

Nonetheless, it may still be the case that E confirms (or disconfirms) H relative to a more restricted set of background knowledge propositions, K', which does not entail E This suggests how E, though already known, might count as confirming evidence for H The idea is simply to subtract E in some suitable way from K and to suggest that such "old evidence" has confirming power when $P(H/E\&K') > P(H/K')$ There are a variety of proposed ways to perform the relevant subtraction Some, so-called "Historical Proposals," appeal to the actual historical epistemic situation of the relevant agent prior to learning that E is the case They hold that E confirms H for S iff prior to S coming to know E, $P(H/E\&K)$ was greater than $P(H/K)$ [5] A more plausible approach is to be found in proposals which appeal to counterfactual (rather than actual historical) epistemic situations in order to solve the problem These, so-called "Counterfactual Proposals," hold that E confirms H for S iff were S to fail to know E, then $P(H/E\&K)$ would be greater than $P(H/K)$ [6]

The adequacy of these proposed solutions to the problem of old evidence has been much debated, but whatever the value of these proposals for accommodating standard cases of old evidence, there is no way *any* such proposal will allow E_1 or E_2 to serve as evidence This is for one simple reason unlike the situation with almost all other cases of old evidence, it is impossible to be in an epistemic situation in which one rationally doubts (that is, lacks certain knowledge of) one's own existence or in which the epistemic probability of one's existence is less than 1 E_1 and E_2 are such that a fully rational agent is necessarily certain of their truth Hence, they are a necessary part of the background knowledge a fully rational being has in any epistemic situation whatsoever and, given PR, cannot serve as evidence E_1 and E_2 thus present us with what I shall call "the strong problem of old evidence "[7]

---

[4] Glymour, *Theory and Evidence* (Princeton University Press, 1980)

[5] Cf Paul Horwich, *Probability and Evidence* (New York Cambridge, 1982)

[6] Cf Colin Howson and Peter Urbach, *Scientific Reasoning The Bayesian Approach* (Chicago Open Court, 1989)

My argument thus far may be precisely stated as follows

(1) E is evidence for H, relative to K, iff $P(H/E\&K) > P(H/K)$ [8]
(2) The probabilities in (1) are epistemic probabilities
(3) Epistemic probabilities are relative to an epistemic situation
(4) The epistemic probability of "I exist" is 1 for me in every possible epistemic situation I could occupy [9]

(5) "I exist" cannot be evidence for me

II DUTCH BOOKS AND CARTESIAN KNOWLEDGE *ALEA LUDO, ERGO SUM*

Subjective Bayesians of a particularly pure sort might reject premise (4) because they maintain that probabilistic coherence is the only necessary constraint on (synchronic) rational credence, and probabilistic coherence does not require that any contingent proposition receive a probability of 1 While the Bayesian theory of rational belief requires that a rational agent's degrees of belief satisfy the axioms of the probability calculus, there are a variety of formulations of those axioms In particular, there are a variety of formulations of the re-

[7] Henceforth, I will drop constant reference to E_1 and E_2 and generally refer only to E_1 I assume that my arguments apply equally to both propositions

[8] A version of this problem arises as well for Likelihoodists (see Anthony W F Edwards, *Likelihood* (New York Cambridge, 1972)), who decline to think of evidence as raising the probability of a theory and claim that E "supports" (in their stipulative sense) H_1 over H_2, relative to K, iff $P(E/H_1\&K) = P(E/H_2\&K)$ Perhaps my argument in this section provides further justification for Elliott Sober's likelihoodist skepticism regarding the possibility of E_1 serving as evidence See his "The Design Argument," in William Mann, ed , *The Blackwell Guide to the Philosophy of Religion* (Malden, MA Blackwell, 2005), pp 117–47 It should be noted that the objection to premise (2) discussed in section III might be more successful if offered by an "objective likelihoodist" as the likelihoodist notion of "support" is not clearly grounded in same intuitive rationale as PR

[9] Accounts of belief on which *all* belief is fundamentally belief *de se* would also seem to imply that one can have no epistemic probabilities without certainty regarding one's existence Furthermore, the intriguing analysis of epistemic probability advanced by Barbara Davidson and Ronald Pargetter, according to which all *de dicto* epistemic probability is *de re* probability relative to a grasp of one's world under the description "world of which I am a part," implies that the epistemic probability of "I exist" is necessarily 1 See their "Probability *De Re*," *The Southern Journal of Philosophy*, XXV (1987) 313–30, and Pargetter's "Conditional Epistemic Probability," *The Southern Journal of Philosophy*, XXVI (1988) 555–71

quirement that a certain class of propositions receive a probability of 1 The standard formulations of that axiom include the following

$P(t) = 1$ if t is a tautology or logical truth[10]

$P(A) = 1$ if A is true in all possible worlds[11]

Let us call the first formulation "the logical formulation" and the second "the modal formulation " A less frequently proposed formulation, which we might call "the epistemic formulation," is

$P(\text{certain proposition}) = 1$[12]

It should be clear that these three formulations are not equivalent First, it is supposed by most philosophers that there are necessary truths which are not logical truths If so, then the logical and modal formulations are not equivalent As well, if rationality requires, as maintained by the epistemic formulation, assigning 1 to a certain proposition, then because the Cartesian propositions seem certain on any reasonable construal of that claim, they must receive a probability of 1, even though this diverges from the requirements of the logical and modal formulations So, if there is any objection to premise (4) from those who hold that only probabilistic coherence is required for rationality, it must be grounded in the logical or modal formulations of the relevant axiom and not in the epistemic formulation

An initial (and, I believe, fully satisfactory) reply to the claim that rationality does not require a probability of 1 for one's existence is simply that it is extremely plausible on intuitive grounds that this is false The *philosophical* defense of axioms of epistemic probability requires argument and the accommodation of our strongest intuitions Hence, if subjectivists deny that rationality requires a probability of 1 for one's existence, they demonstrate the clear poverty of their theory as an account of epistemic probability However, as this objection may be dismissed by some subjectivists with bluster about the unreliability or irrelevance of such intuitions, I shall now argue that the subjectivists' own main argument that rationality requires probabilistic coherence (as standardly formulated), if sound, also justifies imposing a requirement that one have a probability of 1 for one's own existence

---

[10] See Howson and Urbach, *Scientific Reasoning The Bayesian Approach*, Horwich, *Probability and Evidence*

[11] See John Earman, *Bayes or Bust* (Cambridge MIT, 1992), Alvin Plantinga, *Warrant and Proper Function* (New York Oxford, 1993)

[12] See Ian Hacking, *An Introduction to Probability and Inductive Logic* (New York Cambridge, 2001), Roger Rosencrantz, *Foundations and Applications of Inductive Probability* (Atascadero, CA Ridgeview, 1981)

That justification for coherence as a necessary condition of rationality appeals to the thought that in violating the axioms one opens oneself up to a series of bets (a "Dutch book" or "sure loss contract") which one is guaranteed to lose The idea is that there is something defective in the doxastic state of an agent who is such that she is in a position to accept as fair a series of wagers which is such that, no matter how the world turns out, she must lose In the words of Paul Horwich, "if a person is rational, he will distribute his probabilities—his degrees of belief—in accord with [the usual probability axioms] For only if he does this will he be able to avoid a so-called Dutch book being made against him "[13]

A standard motivation for imposing the requirement that an agent assign a probability of 1 to any tautology or logical truth is that such a proposition is guaranteed to "turn out" true and so, insofar as a Dutch book can be made against an agent when her degrees of belief sanction bets which will result in a net loss no matter how the world turns out, any partial belief of degree n which is less than 1 in such a proposition will necessarily result in a loss on the corresponding bet against the proposition based on degree of belief $1-n$ This may be what inclines some to hold that the requirement of maximal credence extends to all necessary truths since any bet against such a truth is bound to be a losing one

However, while it could "turn out" that most contingent propositions are false, it could not turn out, relative to the fact that one is in the betting situation envisaged in the Dutch book arguments, that one not exist If it is the guarantee of losing the corresponding bet that requires one to have a probability of 1 for any necessary truth, then a similar guarantee of loss should require one to assign a probability of 1 to one's own existence because any wager one makes against that proposition is bound to produce a loss If so, then the Dutch book justification for the standard requirement that one have a probability of 1 for any necessary truth will do as well for the requirement that one assign a probability of 1 to one's own existence [14]

This exclusive focus on the necessity of loss might, however, be thought mistaken as I am surely not irrational in being less than fully certain that Water is H_2O or that George and Barbara Bush are the biological parents of George H W Bush, especially if I lack relevant empirical evidence While many hold that these are necessarily true propositions, my rationality is not necessarily impugned by failure to invest full confidence in them This shows, I think, that it cannot be the mere *existence* of a series of loss-guaranteeing bets which are sanctioned by my degrees of belief that is sufficient to convict me of irrationality Rather, the role of the bookie is to make vivid the fact that the book must be, in a certain sense, constructible in a particular epistemic situation Clearly, if the bookie knows some contingent fact which I do not, his ability to make a book against me does not in any way impugn my rationality This is why standard formulations suppose that the bookie knows nothing contingent but the credences of his potential victim Similarly, if the bookie knows some necessary truth the justification of which is partly empirical, such as the candidate *a posteriori* necessary truths mentioned above, his ability to take advantage of my lack of certainty regarding such a proposition does not impugn my rationality

---

[13] Horwich, *Probability and Evidence*, pp 19–20

[14] Similarly, there is a Dutch book argument for the requirement that one have maximal credence in claims about one's credence Howard Sobel, for these reasons, holds that "introspective omniscience is part of an ideal for intellects" in his "Self-Doubts and Dutch Strategies," *Australasian Journal of Philosophy*, LXV (1987) 56–81, p 69

This is nicely captured by David Christensen, when, in a searching exploration of the force behind Dutch Book arguments, he suggests that

> It is plausible that, if a set of betting odds allows someone to devise *a priori* a way of exploiting those odds to inflict a sure loss, then there is something amiss with those betting odds The fact that the diagnosis can be made *a priori* indicates that the defect is not one of fitting the beliefs with the way the world happens to be Like deductive inconsistency, it is a defect internal to the agent's belief system [15]

Christensen's plausible account of the underlying motivation of Dutch Book arguments quite naturally extends to justify requiring a probability of 1 for one's existence First, if one has (partial) beliefs about anything, then one exists and so a failure to fully believe that one exists is clearly a defect "internal to" one's belief system Second, while the world "happens to be" such as to contain one, one's belief about one's existence does not just "happen to be" true and so the defect is not one of *fitting* that belief with the way the world happens to be Finally, it seems that a book against an agent uncertain of her own existence can be made by a bookie with no contingent knowledge of the world beyond knowledge of the existence of the potential victim [16]

---

[15] Christensen, "Dutch-Book Arguments Depragmatized Epistemic Consistency for Partial Believers," this JOURNAL, XCIII, 9 (September 1996) 450–79, p 457

[16] I assume here that the bookie is in a position to grasp (at least indirectly) the relevant proposition which I know in knowing E_I A full evaluation of this issue would require a detailed account of how the bookie is to produce the relevant set of sure loss bets

Perhaps, then, the real force of Dutch book arguments is that they reveal that one has a defect in one's belief system which is, in principle, discoverable solely on the basis of rational reflection, independent of how the nonmental contingent world presently happens to be. This captures the notion that those necessary truths discoverable independent of empirical investigation of the world ought to be accorded a probability of 1. It properly does not, contrary to the modal formulation of the relevant axiom, imply that every necessary truth must receive a probability of 1. However, it implies that one's own existence ought to be accorded the maximum probability.[17] So, any set of axioms defensible on a properly justified Dutch book argument will include the requirement that "I exist" have a probability of 1. I conclude that if rationality requires immunity from a Dutch book, then it requires certainty regarding one's existence. Hence, not only is premise (4) quite obviously true, but those who wish to argue that rationality requires coherence (with its standard axioms) on Dutch book grounds, ought to accept premise (4) on those same grounds.[18]

III. LOGICAL PROBABILITY AND CARTESIAN KNOWLEDGE

For those who agree that PR appeals to an inequality between epistemic probabilities, my conclusion is established. However, some may attempt to undermine premise (2) by claiming that PR refers to *logical* (or *inductive*) probabilities rather than *epistemic* probabilities. Actually avoiding the conclusion of my argument would also require maintaining that these probabilities exist even where no possible epistemic situation consists only of the propositions in *K* and so *K* could not be the total background knowledge of any possible rational agent. Such proponents of logical or inductive probability may respond to my arguments by suggesting that my existence can be positive relevance evidence for a hypothesis, provided that we take PR to appeal to such nonepistemic probabilities.

A possible basis for such an "Objective Bayesian" rejoinder to my argument can be found in the work of inductive logicians like Rudolf Carnap and Brian Skyrms, and has been developed in "two-function"

---

[17] It is, of course, a matter of some debate whether to classify one's knowledge of E_1 as *a priori* or *a posteriori*. The difficulty is that knowledge of E_1 seems not to depend on experience but it is also not justified by whatever nonexperiential source one takes to justify us in believing *a priori* necessary truths.

[18] Some probabilists avoid Dutch book arguments and appeal instead to the notion that one minimizes the inaccuracy of one's beliefs only when one's beliefs are probabilistically coherent (see James Joyce, "A Nonpragmatic Vindication of Probabilism," *Philosophy of Science*, LXV (1998): 575–603). However, if it is the accuracy of *one's own* partial beliefs that should be maximized, then surely failure to assign a probability of 1 to E_1 will be less accurate than would be an assignment of 1.

Bayesian approaches advanced by James Hawthorne and Mark Lange,[19] each of whom is explicitly attempting to avoid the problem of old evidence On such approaches, a rational agent is characterized by *two* probability functions The first is a function representing an agent's current *rational degree-of-belief* in each proposition This function is the one that typically changes over time with the receipt of new information, as rational degree of belief (or epistemic probability) is relative to one's evidence The second function represents the *degree-of-support* which a proposition provides another proposition The exact nature of this second function is a matter of disagreement among these objective Bayesians, but they all agree that it will usually vary from an agent's degree-of-belief function[20]

This last feature is what allows for the solution to the standard problem of old evidence While it may be true that, at any time when an agent has certain knowledge of E, her degree-of-belief function will contain $P(E) = 1$, it will not, allegedly, when E is contingent, be true that her degree-of-support function will contain $P(E) = 1$ As Hawthorne puts the point, the support function will not assign E support-strength 1 unless all of the relevant alternative hypotheses yield degree-1 likelihoods for E (*op cit*, p 313) Hence, so long as the support function includes $P(E_1)$ of less than 1, the problem I have been developing would seem to be avoided and the two-function account will allow E_1 to serve as evidence, even given PR

This sort of Bayesianism and its general strategy for dealing with the standard problem of old evidence seem to me extremely plausible However, it cannot avoid the *strong* problem of old evidence raised by E_1 This is because, as noted above, while almost every other contingent putative evidence statement for which a possible old evidence problem might arise is a statement which a rational agent could fail to know, the same is not true for E_1 Hence, as I shall now show, the version of the problem of old evidence raised by Cartesian knowledge is so strong that it cannot be overcome without destroying the required connection between logical and epistemic probabilities

---

[19] See Carnap, *The Logical Foundations of Probability*, and "Inductive Logic and Rational Decisions," in Carnap and Richard Jeffrey, eds , *Studies in Inductive Logic and Probability I* (Berkeley California UP, 1971), pp 5–31, Skyrms, *Choice and Chance*, 3rd edition (Belmont, CA Wadsworth, 1986), Hawthorne, "*Degree-of-Belief* and *Degree-of-Support* Why Bayesians Need Both Notions," *Mind*, CXIV (2005) 277–320, Lange, "Calibration and the Epistemological Role of Bayesian Conditionalization," this JOURNAL, XCVI, 6 (June 1999) 294–324

[20] "Degree-of-belief" and "degree-of-support" are Hawthorne's terms He disavows the Carnapian claim that prior probabilities can be fixed on purely logical or syntactic grounds, but he emphasizes the "objective" and "quasi-logical" character of many likelihoods in the support function Carnap calls the first function a "rational credence function" and the second a "credibility function "

Rather than attempting to engage here with the details and nuances of variations on this view, I shall provide an argument directed against their common core My argument consists of a dilemma for any proponent of such a two-function view who wishes to evade the argument of section I Either the envisaged support function probabilities are *possible* epistemic probabilities or they are not If they are possible epistemic probabilities, then my arguments in sections I and II would seem to show that the appeal to such logical or inductive probabilities will be unavailing in avoiding my conclusion [21] If they are not possible epistemic probabilities, then while the inequality required by PR may obtain with respect to the support function, the connection between PR and its intuitive rationale will have been severed Hence, no appeal to logical probability can allow knowledge of E_1 to serve as evidence in the sense which PR attempts to capture

Taking the first horn as established by the arguments of earlier sections, what can be said for the second horn? If support function probabilities are not possible epistemic probabilities, then it may be that $P(E_1)$ is less than 1 and it may be that for some H, $P(H/K\&E_1) > P(H/K)$, as required in order to satisfy PR The resulting account would, however, allow for the satisfaction of PR in a case where the fundamental intuitive rationale for PR is violated That rationale, recall, is that E is evidence for H (relative to K) if and only if knowledge of E would make it rational to be more confident of H [22] The comparative "more" clearly requires that we compare a situation in which one does not know E with one in which one does know E However, if E is something one could not fail to know with certainty, then such a comparison will be impossible and E cannot satisfy the rationale at issue So, while the proposed appeal to support functions nicely avoids the *standard* problem of old evidence while respecting the intuitive rationale behind PR, it yields an account of evidence that cannot solve the *strong* problem of old evidence presented by E_1 without abandoning the rationale behind PR

---

[21] Two-function Bayesians are not always entirely clear what constraints are placed on support functions (beyond adherence to the usual axioms) and why However, some logical probabilists make use of Dutch book arguments in arguing for the correctness of a favored support function See, for example, Carnap, *The Logical Foundations of Probability*, John Kemeny, "Carnap's Theory of Probability and Induction," in Paul Schilpp, ed , *The Philosophy of Rudolf Carnap* (La Salle, IL Open Court, 1963), pp 711–38 If my argument in section II is sound, however, adherence to such arguments requires a probability of 1 for E_1 in the support function Hawthorne leaves open the possibility that a support function is to be identified with some suitably chosen (counterfactual) rational degree-of-belief function That route clearly also forecloses the use of E_1 or E_2 as evidence

[22] See Patrick Maher, "Subjective and Objective Confirmation," *Philosophy of Science*, LXIII (1996) 149–74

The difficulty E_1 presents for two-function Bayesians may be clarified by further attention to the connection such theorists wish to forge between inductive probability and epistemic probability As Hawthorne notes, the "primary role of a support function in the two-function model is to underwrite and justify the agent's belief strengths for hypotheses" (*op cit*, p 309) Given that, "intuitively, each agent's belief strength for a hypothesis should be just the degree to which the hypothesis is supported, according to her support function, by the totality of the evidence she has acquired" together with whatever fixes one's prior probabilities (*op cit*, p 307), Hawthorne holds that the fundamental connection between the two functions is that a rational agent ought, at each time *t*, to "align" his degree-of-belief function with his degree-of-support function conditional on his total evidence at *t*

As I have already allowed, this account provides a way around the standard problem of old evidence and it does so in a way that *generally* accommodates what I have claimed is the guiding intuition behind the positive relevance criterion of evidence—that *E* is evidence for *H* (relative to *K*) if and only if knowledge of *E* would make it rational to be more confident of *H* It does so because, given the claim that a rational agent aligns his degree-of-belief function with his degree-of-support function, a rational agent with background knowledge *K* would, in the absence of knowledge of *E*, have a lower degree-of-belief in *H* than he would when he knows *E* just when his support function includes $P(H/E\&K) > P(H/K)$ Exactly here, however, we find an appeal to possible epistemic situations and if *E* is something one *could not fail* to know, then either the possibility of satisfying PR remains because *E* has a support-function probability of less than 1 but alignment with both sides of PR will be impossible, or the support function admits a probability of 1 for *E* and preserves the possibility of alignment with both sides of PR while rendering impossible the inequality it requires if *E* is to serve as evidence [23]

---

[23] Lange also touches on some of these issues when he notes that his account begins with an initial probability function (much like Hawthorne's "degree-of-support function" in being distinct from one's current rational degrees of belief) and he suggests one question regarding that initial function is why it constitutes "a reasonable place to begin one's doxastic life" (p 303, fn 19) If that is the question, then it seems clear that the probability function at issue must include a probability of 1 for one's own existence Furthermore, as Lange notes that his envisaged initial probability distribution is "like a counterfactual degree of belief in that it may be a fictitious state of opinion," if the prior probability distribution must at least be a *possible* state of opinion, it would seem to require a probability of 1 for one's existence

This fundamental conflict between PR (when construed as a claim about support function probabilities) and the alignment requirement is simply elided when it is supposed that every contingent proposition must be capable of serving as evidence and that an ideal rational agent "begins"[24] or "could" under "idealized conditions" have begun[25] with a rational degree of belief function which is identical to his support function conditional on a tautology [26] Once this conflict is brought to light, it is clear that the envisaged fully general connection between logical and epistemic probabilities cannot be sustained in a manner according with PR It seems, then, that such proponents of inductive or logical probability cannot escape the conclusion of my argument in section I by denying premise (2) as that premise must be true if PR is to remain properly grounded in its intuitive rationale

Before proceeding to consider the further implications of my central theses, we might pause to note that the result before us is, when considered in an appropriate light, quite plausible Removing our Bayesian lenses, if we accept that one intuitive notion of evidence is a notion of knowledge which does or could justify us in changing our credence in some hypothesis, we should find it quite plausible that knowledge which is an essential part of any possible epistemic situation in which we could find ourselves cannot serve as that kind of evidence Such knowledge, if we are perfectly rational, could never have come newly to our attention and require adjustment in our view of the world in order to accommodate it An essential part of any possible epistemic situation cannot, under any circumstances, provide a fully rational creature with any reason to change her view

IV GOD, THE MULTIVERSE AND DOOM

If my arguments above are successful, then E_1 and E_2 cannot serve as evidence for us as they are necessary components of any possible epistemic situation which we might occupy Furthermore, propositions known *a priori* to be entailed by E_1 and E_2 cannot count as evidence as they must also have a probability of 1 Such propositions include the existential generalizations of E_1 and E_2

---

[24] Hawthorne, p 307

[25] Carnap, "Inductive Logic and Rational Decisions," p 18

[26] I am assuming throughout this paper that "I exist" is not a tautology or logical truth David Kaplan suggests that it is a logical truth of the "logic of demonstratives " See his "On the Logic of Demonstratives," *Journal of Philosophical Logic*, VIII (1978) 81–98 If the concept of a tautology or of a logical truth is stretched so as to include truths in the logic of demonstratives, then, of course, E_1 presents no obstacle to the notion of epistemic probabilities relative to purely logical knowledge However, it is also then quite clear that the logical probability of E_1 is 1 and so E_1 cannot serve as evidence on the logical conception of probability

E_3 = Something exists

E_4 = A conscious being exists

Furthermore, the arguments above justify the claim that *E* can serve as evidence for *S*, relative to background knowledge *K*, only if it is logically possible that $P(E/K) < 1$ The remaining task is to demonstrate how these results undermine some standard formulations of a number of significant contemporary probabilistic philosophical arguments

IV A Arguments for Theism Philosophers of religion have shown an increasing interest in probabilistic, rather than the traditional deductive, arguments for and against the existence of God The most systematic probabilistic case for the existence of the traditional God has been put forward by Richard Swinburne [27] Swinburne explicitly attempts to demonstrate that various known facts are evidence, in exactly the sense formalized by PR, for the existence of the theistic God On Swinburne's view, the simplest probabilistic version of the cosmological argument is based on an evidence statement which merely asserts the existence of a finite object Where *H* is the theistic hypothesis, he claims that "if we take *E*, most generally, as the existence of a finite object $P(E/H\&K)$ will exceed $P(E/{\sim}H\&K)$" (*ibid*, p 131) and so the existence of a finite object raises the probability of God's existence and constitutes evidence in the sense of PR for God's existence As Swinburne's *E* is, given the knowledge that I am finite,[28] entailed by E_1, my earlier arguments demonstrate that this probabilistic recasting of the cosmological argument fails to provide any evidence for the existence of God [29]

---

[27] Swinburne, *The Existence of God*, 2nd edition (New York Oxford, 1990)

[28] Given the conception of finitude at issue, this knowledge is surely also available merely on reflection as one can know on mere reflection that one is not omniscient (Alternatively, given that E_1 is part of *K*, if ${\sim}H\&K$ is consistent, then it implies the existence of a finite object and $P(E/H\&K)$ could not exceed $P(E/{\sim}H\&K)$) Probabilistic cosmological arguments based upon the existence of a contingent object are similarly vitiated by my arguments if one can tell (as it seems one can) on reflection alone that one is contingent

[29] Given standard Bayesian idealizations, the general principles on which my argument regarding E_1 and E_2 is based also show that neither

* I know *a priori* necessary truths,

nor

** I am a rational being,

nor their *a priori* entailments can serve as evidence for or against any hypothesis, contrary to what some theistic philosophers seem to maintain

Like a number of other philosophers, Swinburne also argues that the existence of conscious persons provides positive relevance evidence for the existence of God (*op cit*, p 174) However, the mere existence of conscious persons cannot, given my earlier results, serve as evidence for any hypothesis whatsoever The issues here are complicated by the fact that Swinburne sometimes appears to take the relevant evidence to be the existence of *embodied* conscious agents Nonetheless, as I have shown that certainty regarding the existence of conscious creatures is a necessary component of any epistemic situation, the proponent of such an argument from embodiment must hold that there is a possible epistemic situation in which one lacks certainty regarding the existence of embodied conscious creatures and that one's discovering such embodiment is more likely on theism than on its negation Given merely certainty that one exists as a conscious creature (whether embodied or not), I see no reason to think one is more likely to be embodied given theism than given its negation Clearly, at the very least, the theist must give up on any argument from the existence of conscious creatures and make a case for the positive relevance of embodiment, relative to a situation in which one is certain that one exists *qua* conscious creature

IV B Arguments from Fine-Tuning Recent discoveries in fundamental cosmology have suggested to many scientists and philosophers that our universe is "fine-tuned for life " Put more precisely, such writers claim that the value of various variables in the fundamental account of the universe do not follow from our fundamental theory [30] Furthermore, for many such variables, the range of values which allow for the existence of life is extremely small, relative to the range of theoretically possible values It is alleged that, taking as background knowledge the aforementioned claims, the fact that the relevant variables actually have life-permitting values is positive relevance evidence for at least one of two hypotheses The first is the hypothesis of traditional theism [31] The second is the hypothesis that there actually exist many physical universes (a "Multiverse"), of which ours is merely one [32] The first hypothesis is said to render the existence of life more probable than does atheism, given some axiological facts and the nature of God The second hypothesis is said to render the existence of life more probable than does the hypothesis of a single universe because it provides more "chances" for a life-permitting universe

---

[30] For skepticism about such arguments, see Tim McGrew, Lydia McGrew, and Eric Vestrup, "Probabilities and the Fine-Tuning Argument A Sceptical View," *Mind*, cx (2001) 1027–38

[31] See Swinburne, *The Existence of God*

[32] See Rodney Holder, "Fine-Tuning, Multiple Universes and Theism," *Noûs*, xxxvi (2002) 295–312, Peter van Inwagen, *Metaphysics* (Boulder, CO Westview, 1993)

While I have just framed the argument in terms of the fact of the universe being "life-permitting," it is clear that most proponents of the argument are actually focused on the existence of *conscious* living creatures This is clearest in the theistic version of the argument as God is presumed to value the existence of conscious creatures and not the mere existence of unconscious biological life as such Furthermore, most proponents of the argument clearly assume that we are living physical beings or are suitably dependent on the existence of such beings for our existence as conscious creatures Indeed, the usual assumption is that *we* would not exist if the relevant constants had different values Hence, the arguments of this paper undermine certain simple formulations of the reasoning just rehearsed This is because, relative to knowledge that one is part of a universe, or that one is or is dependent upon a particular living being, and the usual assumption that certain "fine-tuning" conditions must be met for conscious living creatures to exist in a universe, the probability that the universe one is in meets those conditions necessary for conscious life is 1, and so the fact that one occupies such a universe or that such a universe exists cannot serve as evidence for any hypothesis whatsoever Hence, it appears that, given the usual shared assumption noted above, both theists and proponents of a multiverse are mistaken in thinking that the aforementioned cosmological discoveries provide any evidence for their respective hypotheses[33]

It seems appropriate to note that we have here a novel vindication of a nontrivial version of the "Anthropic Principle "[34] There are many different versions of this infamous principle[35] Some hold that our existence requires no explanation or that our existence (and the conditions which must obtain for our existence) are somehow inevitable Such claims seem to me at best dubious However, my arguments do provide a clear Bayesian grounding of the claim that knowledge of the obtaining of any condition which is known (or assumed on all the relevant competing hypotheses) to be necessary for our own conscious existence cannot serve as confirming evidence, just as some versions of the Anthropic Principle (including the one most relevant to responding to the fine-tuning argument) appear to maintain.[36]

---

[33] For a much more detailed discussion of more precisely formulated fine-tuning arguments and an attempt to reduce those arguments to arguments based on E_1 or E_2, see Kai Draper, Paul Draper, and Joel Pust, "Probabilistic Arguments for Multiple Universes," *Pacific Philosophical Quarterly* (forthcoming) That paper also adumbrates parts of the argument of section I

[34] John Earman notes, in his "The SAP Also Rises A Critical Examination of the Anthropic Principle," *American Philosophical Quarterly*, XXIV (1987) 307–17, that some versions of the principle are merely trivial instances of the multiplication axiom of probability

[35] See John D Barrow and Frank J Tipler, *The Anthropic Cosmological Principle* (New York Oxford, 1986), Brandon Carter, "The Anthropic Principle and Large Number Coincidences," in Malcolm Longair, ed , *Confrontation of Cosmological Theories with Observation* (Boston Reidel, 1974), pp 291–98

IV. C. The Doomsday Argument. One of the most interesting and controversial recent philosophical applications of Bayesian ideas is to be found in what has come to be known as "The Doomsday Argument." Though the argument can be formulated in a variety of ways, the core idea of the argument is that information about one's birth rank in the human species provides more confirmation (that is, relative confirmation) of the claim that there will be fewer total human beings in the history of the universe (the "Doom Soon" hypothesis) than it does less pessimistic hypotheses ("Doom Later" hypotheses).[37] The modest and highly simplified version of the argument presented by Darren Bradley and Brandon Fitelson[38] asks that we imagine three possibilities for how many people there are, and will ever be, in the entire universe. According to H_1, there will be one person, according to H_2, there will be two people, and according to H_3, three people. According to this version of the Doomsday Argument, learning that you are the second person to exist in the universe is, given natural assumptions, confirming evidence for H_2. More generally, one's birth rank provides the most confirmation for that hypothesis (consistent with one's birth rank) on which there are the fewest total people.

Whatever the merits of the Doomsday Argument in this or its other formulations, the theses of this paper undermine a common rejoinder to the argument. That rejoinder accepts the above reasoning for a probability increase in the Doom Soon hypothesis, but maintains that we have additional evidence which, when antecedently taken account of, exactly counterbalances the aforementioned probability increase, leaving us with precisely whatever probabilities the various hypotheses regarding the duration of the human species have independent of this alleged evidence The additional evidence which is alleged to provide this convenient "cancellation" of the Doomsday Argument is our own existence[39] Among the proponents of this rejoinder, Ken Olum claims that "in the scenario where the human race is very long-lived and there are many humans altogether, there is a greater 'chance to exist at all' than in the scenario where the race is soon to die out" (*op cit*, p 166) and Paul Bartha and Christopher Hitchcock claim that the doomsday argument is flawed because "our being alive at all does have evidential bearing upon the number of people that will ever live" (*op cit*, p 13) The basic idea of this reply is that our existence provides evidence that there will be many more persons and our existence with a given birth rank provides evidence that there will be fewer persons So, as Bartha and Hitchcock note, "in the simplest cases, these two effects cancel one another out, yielding the happy result that our existence now tells us nothing whatsoever about the coming of Doom" (*op cit*, p 18)

---

[36] Notice that the objection I have raised cannot be evaded by arguments utilizing an evidence statement such as "Other conscious beings exist" or "More than one finite being exists" because, given that the relevant Cartesian knowledge is included in the epistemic situation, the relevant conditional probabilities will not clearly differ in the way required.

[37] The argument's title is perhaps immodest as some versions allow that the less pessimistic hypotheses may remain vastly more probable than the more pessimistic ones.

[38] Bradley and Fitelson, "Monty Hall, Doomsday and Confirmation," *Analysis*, LXIII (2003): 23–31.

If I am correct, this rejoinder to the Doomsday Argument fails as it explicitly treats knowledge of one's existence as evidence in the sense captured by PR and so runs afoul of the strong problem of old evidence Bartha and Hitchcock admit that some will look askance at their proposal because "the prior probabilities are ones we literally could never have had" and "we never actually have subjective probabilities in which our existence is not taken as a given " Claiming that the problem here is just another instance of the problem of old evidence, they tentatively endorse an appeal to hypothetical probability distributions as a means of solving that problem While I am quite sympathetic to such appeals to solve most instances of the problem of old evidence, I believe that they have failed to appreciate the fact that the hypothetical probability distribution that their rejoinder requires is, unlike that required for solving standard instances of the problem of old evidence, one that is logically impossible Their reply fails because we could not have (coherent or rational) subjective probabilities in which our existence is not taken as a given So, if the Doomsday Argument is to be defeated, it must be by some other means[40]

---

[39] See Dennis Dieks, "Doomsday—Or The Dangers of Statistics," *Philosophical Quarterly*, XLII (1992) 78–84, Ken Olum, "The Doomsday Argument and the Number of Possible Observers," *Philosophical Quarterly*, LII (2002) 164-184, Paul Bartha and Christopher Hitchcock, "No One Knows the Date or the Hour An Unorthodox Application of Rev Bayes's Theorem," *Philosophy of Science* (Proceedings), LXVI (1999) S229–S353

[40] For an account of how some of these ideas create difficulties for one recent attempt to solve the Sleeping Beauty problem, see my "Horgan on Sleeping Beauty," *Synthese* (forthcoming)

V ESCAPING THE FIRING SQUAD

It may seem to some that the position defended thus far is open to clear counterexamples A putative counterexample runs as follows

> On a certain occasion the firing squad aim their rifles at the prisoner to be executed There are twelve expert marksmen in the firing squad, and they fire twelve rounds each However, on this occasion all 144 shots miss The prisoner laughs and comments that the event is not something requiring any explanation because if the marksmen had not missed, he would not be here to observe them having done so But of course, the prisoner's comment is absurd, the marksmen all having missed is indeed something requiring explanation, and so too is what goes with it—the prisoner's being alive to observe it And the explanation will be either that it was an accident (a most unusual chance event) or that it was planned (such as, all the marksmen had been bribed to miss) [41]

Setting aside the issue of explanation, it might be suggested that the view I have been advancing is incapable of accommodating the widespread intuition that the prisoner has evidence (after the shots are fired) for the hypothesis that the marksmen intended to miss (H_1) and against the hypothesis that they intended to kill (H_2)

It is reasonable to hold that the prisoner has *some* evidence for H_1, at least because standard presentations of the case leave crucial details unclear For example, his knowledge that

E_5 No bullet hit me

is evidence for H_1 [42] However, this is perfectly consistent with my position, as his *existence* is not here alleged to be evidence In order to render the case a potential problem for my thesis, we must attempt to imagine a case in which the only thing that the prisoner comes to know is that he exists at the appropriate time

Consider, then, a variant case in which the prisoner is entirely deprived of all sensory input from the moment after the firing of the bullets In this case, it may seem that the prisoner is aware only of his continued conscious existence and so if he has evidence, it must consist of Cartesian knowledge such as

E_6 I exist now

---

[41] Swinburne, *The Existence of God*, p 171, see also p 138, and van Inwagen, *Metaphysics*, p 135, for structurally identical examples produced in response to objections based on some version of the Anthropic Principle

[42] Sober, "The Design Argument "

Given my arguments, E_6 can serve as evidence only if there is some possible epistemic situation relative to which E_6 has a probability of less than 1

Although the knowledge in question involves the vexing indexical "now," even without a detailed investigation of such knowledge we can distinguish views on which token instances of E_6 thought (and known) at distinct moments in time express distinct propositions and views on which they do not On the latter views, it seems that instances of E_6 must have, relative to any possible epistemic situation, a probability of 1 and, hence, no instance of E_6 can serve as evidence However, before it is concluded that the modified Firing Squad case would, on such views, present us with a potential counterexample to my claims, it should be noted that, in spite of our attempt to imagine otherwise, the prisoner clearly has other new knowledge which, for all I have argued, can serve as evidence for H_1 In the prisoner's case, it seems that propositions such as

E_7 I exist after the firing squad fires

or, given certain assumptions about the date and time of the possible execution,

E_8 I exist at 4 30 P M on January 15th, 2005

can serve as evidence[43] So, even *if* we hold that all instances of E_6 express knowledge of the same proposition, we have in the Firing Squad case no reason to question the arguments of this paper

Though knowledge of E_7 and E_8 is also available to serve as evidence on views according to which instances of E_6 express knowledge of *different* propositions, on such views it remains possible that one might occupy an epistemic position relative to which a given instance of E_6 is not known (even though another instance, expressing a distinct proposition, would be known with certainty) On such views it may be possible to construe the new indexical knowledge as evidence Even *if*, however, such an account can be made to work, it is crucial to note that the temporally indexical Cartesian proposition, expressed by an instance of E_6, which might thereby serve as evidence in the Firing Squad case can do so *only relative to a possible epistemic situation which must include certainty regarding one's (then) current existence* In other words, while one's current (at *t*) existence might, on these views, serve

---

[43] It might be objected that E_7 is equivalent to the conjunction of E_6 and

*** Now is after the firing squad fired

and that E_7 cannot be evidence if E_6 and *** are not But while E_6&*** entails E_7, the converse does not hold and they are not equivalent as positive relevance evidence, this can be so only if knowledge of one's current existence makes it rational to have greater credence in some hypothesis than would be rational if one were not to know of one's current (at *t*) existence Such a situation, however, would be one in which one was certain of one's current (at *t*–1) existence Hence, even if we disregard knowledge of propositions like E_7 and E_8, it can, at best, be said not that one's existence at a time counts as evidence, but that one's *continued* existence may count as evidence These are the crucial facts missed by those who think cases like the Firing Squad establish the *general* acceptability of appeals to one's existence as evidence and such facts undermine their attempt to demonstrate that the mere fact of one's existence or consciousness may serve as evidence

My thesis, suitably qualified to accommodate the various possible ways of responding to cases like the Firing Squad, is that (1) instances of the atemporal E_1 clearly cannot serve as evidence because they are an essential part of any epistemic situation and (2) instances of the temporally indexical E_6 are either (a) each expressions of the same proposition and cannot serve as evidence because they are an essential part of any epistemic situation, or (b) each expressions of different propositions which can serve as evidence only by amounting, in virtue of their relation to the relevant contrasting epistemic situation (in which they do not feature but which must itself include knowledge of another instance of E_6), to the proposition that one has continued to exist In this last case, however, the hypothesis in question would have to be relevant to one's continued existence, given that one antecedently existed

It might reasonably be wondered whether this last view, if it turns out to be defensible, would allow reformulation of the various arguments which were shown in section IV to encounter difficulties when based upon E_1 or E_2 Notice, however, that such reformulations would, in essence, concede my earlier claim that E_1 and E_2 cannot serve as evidence and would attempt to formulate different arguments based upon a different evidence statement like E_6 or E_7 Recognizing that much more might be said regarding possible reformulations of the problematic arguments, it is quite difficult to see how such arguments might succeed For instance, it appears that there is little reason to think that one's continued existence or consciousness at some particular time is, given that one does exist *qua* conscious creature, more probable on theism than on its denial Hence, the cosmological argument and the argument from consciousness cannot be easily reformulated with an acceptable evidence statement Likewise, it appears that, given my existence now, my continued existence is no more probable no matter how many universes there might now or ever be and, given my existence in a fine-tuned world, my continued existence in a fine-tuned world is no more probable on theism than its denial Finally, the "cancellation rejoinder" to the Doomsday Argument cannot be reformulated to focus on my knowledge of my continued existence as its focus is on the alleged evidential bearing of my existing at all and, given that I now exist, it seems my continued existence is not differentially probable conditional on the various hypotheses regarding how many people will ever exist So, it seems that our earlier negative results survive the Firing Squad and a difficult burden rests with those who wish to successfully reformulate the earlier arguments

VI CONCLUSION

I have not, of course, shown that E_1 and its relatives cannot serve as evidence on an account of evidence suitably severed from appeal to standard Bayesian foundations [44] Nonetheless, I have shown that, contrary to a number of philosophers, given PR, they cannot serve as evidence and cannot be excluded from the background knowledge relevant to assessing whether knowledge of any proposition is evidence for any hypothesis The strong problem of old evidence demonstrates that those convinced that their existence or some proposition entailed by their existence and relevant background knowledge might be evidence cannot be Bayesians of any sort More importantly, it shows that we cannot treat our existence (or its relevant entailments) as evidence on any conception of evidence grounded in the fundamental intuitive motivation for the positive relevance account We have, it seems clear, a choice between (a) rejecting the notion that the mere fact of one's existence may serve as relevance evidence in the sense captured by PR, and with it a number of apparently important recent philosophical uses of such alleged evidence, or (b) rejecting PR and the plausible intuitive rationale in which it is grounded While I favor the former option, I have here tried to demonstrate that the choice is quite significant and that choose we must

JOEL PUST

University of Delaware

---

[44] In addition, for all I have argued here, it may still be maintained that E_1 is evidence in the *absolute* sense of confirmation, according to which E is evidence for H if and only if $P(H/E)$ is greater than some fixed $n > 5$ This absolute sense involves no clear comparison to an epistemic situation in which one is bereft of knowledge of E The interest of this sense of confirmation is perhaps diminished by noting that it allows a tautology to confirm an empirical hypothesis, a logical truth to confirm itself, and all the contingent conjuncts of E to reduce the probability of H even though E confirms H Furthermore, to my knowledge, no author has claimed that E_1 absolutely confirms any of the hypotheses under discussion in this paper

THE FALSE DICHOTOMY BETWEEN COHERENTISM AND FOUNDATIONALISM*

The distinction between coherentism and foundationalism has a central role both in epistemology and in moral philosophy. However, insufficient attention has been paid to the meaning of the two concepts. Exactly what is required for a set of beliefs or a set of moral judgments to have a coherentist, respectively foundationalist, structure? It is the purpose of the present contribution to provide a tentative answer to these questions in terms of the patterns of support between the elements of the set. In section I, a formal framework for the analysis is provided. In section II, it is applied to an analysis of coherentism and, in section III, to a corresponding analysis of foundationalism. In section IV, support patterns are discussed that deviate from the standard dichotomy between coherentism and foundationalism.

I. FRAMEWORK FOR THE ANALYSIS

Coherentism and foundationalism will be assumed to refer to the structure of support among a set E of discrete objects. Lower-case letters (a, b, ..., x, y, ...) denote elements of E. For the general argument, no assumptions will be made about the nature of these elements. In some of the examples, they will be taken to be representable as sentences in a language. Sentences are our best general-purpose representation both of beliefs and of moral judgments.

According to a classic formulation, coherentism means that "a body of knowledge is a free-floating raft every plank of which helps directly or indirectly to keep all the others in place, and no plank of which would retain its status with no help from the others." In contrast, foundationalism means that "every piece of knowledge stands at the apex of a pyramid that rests on stable and secure foundations whose stability and security does not derive from the upper stories or sections."[1] Hence both coherentism and foundationalism concern a relation of support that holds between elements of E. This support relation will be denoted S. It can be read "supports." In our

---

* I would like to thank Erik J. Olsson for valuable comments on an earlier version of this paper.

[1] Ernest Sosa, "The Raft and the Pyramid: Coherence versus Foundations in the Theory of Knowledge," *Midwest Studies in Philosophy* v, 1 (1980): 3–25, p. 24.

0022-362X/07/0406/290–300

general argument its nature will be left open It can for instance be inferential, explanatory, justificatory, or probabilistic Coherentism and foundationalism will be treated as views on the properties of the pair $\langle E,S \rangle$

Quantifiers will refer to the domain E unless otherwise specified, thus $\forall x$ and $\exists x$ are abbreviations of $\forall x \in E$ and $\exists x \in E$ The ancestral of S will be denoted S^* Hence, aS^*b holds if and only if either aSb or there is a finite series of elements x_1, x_n such that aSx_1, x_kSx_{k+1} for all $1 \leq k < n$, and x_nSb

To the extent that an element of E supports itself, this is a rather trivial form of support that differs from how it can support other elements For simplicity, S will therefore be taken to be irreflexive, that is, we assume that $(\forall x)\neg(xSx)$ [2]

This framework for support relations in a set E is idealized in at least three important ways First, the relation of support is taken to be a relation between pairs of elements of E, not between pairs of subsets of E It is for instance feasible that $\{a,b\}$ supports c although neither a nor b alone supports c, but such a pattern is not representable in the present framework [3] Second, it is assumed that the relevant relationships between elements of E are captured by a relation of support This means that we discuss these relationships only in terms of "forces of attraction," thus excluding "forces of repulsion" (conflicts) that may hold between the elements Third, support is treated as an all-or-nothing concept (and so, essentially, will coherence and foundational structure be treated in what follows) Both in epistemic and ethical applications it is reasonable to assume that support can come in degrees [4] These idealizations are all made in order to construct a simple framework in which the basic nature of coherence and foundational structure can be studied with as few complications as possible

---

[2] There are at least two alternative approaches (1) We can let S be reflexive, and separate out its irreflexive part Nothing of importance will be gained with this somewhat more complex formal structure (2) We can let xSx hold for some but not all $x \in E$, namely for those elements that are taken to be self-supporting This is a reasonable approach, but it will not be investigated here

[3] It is representable in the framework used in Sven Ove Hansson and Erik Olsson, "Providing Foundations for Coherentism," *Erkenntnis*, LI, 2–3 (1999) 243–65 The present framework can be generalized by letting S be a relation on elements of $\mathcal{P}(E)$ rather than E

[4] I have argued elsewhere that in a fully realistic analysis of coherence, it should be taken to come in degrees See my "Coherence in Epistemology and Belief Revision," *Philosophical Studies*, CXXVIII, 1 (2006) 93–108 A similar argument can be made considering foundationalism

II WHAT IS COHERENTISM?

As a first approximation coherentism may be taken to refer to the pattern of support in which each element of E supports all the other elements of E

Universal direct support
$(\forall x)(\forall y)(xSy)$

This is a very strong condition In a large and diverse structure, such as a system of human beliefs or moral judgments, we expect some elements to lack direct connections, even if the structure as a whole is tightly knit together

The implausibility of *Universal direct support* is confirmed in the following sentential example Let E be a logically closed set of sentences, and let S be a support relation that represents deductive support between elements of E A reasonable definition of deductive support should be such that if x deductively supports y, then there is at least one deduction of y that depends on x, hence there should be some z such that $z \nvdash y$ but $x\&z \vdash y$ If we leave open whether or not tautologies can support other sentences and also whether or not they can be supported by them, this leads to the following requirement on a support relation that operates on a set of sentences

Deductive relevance
If xSy then either $\vdash x$, or $\vdash y$ or there is some z such that $z \nvdash y$ but $x\&z \vdash y$

OBSERVATION 1 (1) Let E be a set of sentences and let S satisfy *Universal direct support* and *Deductive relevance* Then it holds for all sentences x and y that if $x \in E \setminus \mathrm{Cn}(\emptyset)$ then there is no element v of $E \setminus \mathrm{Cn}(\emptyset)$ such that v is equivalent with $x \rightarrow y$

(2) Let E be a logically closed set of sentences that contains at least three nonequivalent sentences, and let S satisfy *Deductive relevance* Then it does not satisfy *Universal direct support*

(For proofs of this and the following observations, the reader is referred to the Appendix)

Sosa's definition of coherentism, as quoted above, indicates how a more reasonable definition of coherentism can be constructed every plank should help "directly or indirectly" to keep the others in place We can express this as follows

Universal indirect support
$(\forall x)(\forall y)(xS^{*}y)$

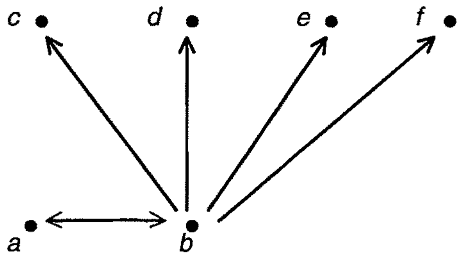


Figure 1 In this and the following diagrams, the dots represent the elements of E, and arrows represent the support relation S in such way that xSy holds if and only if there is an arrow from x to y

For the reasons already given, this is a much more plausible condition Furthermore, the problem for deductively relevant support relations that was exhibited in Observation 1 does not exist if *Universal support* is relaxed to *Universal indirect support*

OBSERVATION 2 Let E be a logically closed set, and let S be the minimal relation that satisfies *Deductive relevance*, that is, xSy if and only if either $\vdash x$, or $\vdash y$ or there is some z such that $z \nvdash y$ but $x \& z \vdash y$ Then $(\forall x)(\forall y)(xS^{*}y)$

I will take *Universal indirect support* as a first approximation of how the standard view of coherentism can be expressed in the present formalism It can, however, reasonably be argued that this condition is yet too strong When writing this, I believe that there is tea in the cup in front of me, and I also believe that some horses allow humans to ride on them It is difficult to conceive of a plausible way in which one of these two beliefs supports the other (They are arguably both supported by the same general beliefs about the veracity of sense impressions, but that does not make them support each other even indirectly) Should my belief system, for this reason, be considered not to be coherent? Examples like this give us reason to search for weaker conditions for the interconnectedness required of a coherent structure The following is a plausible weakening of *Universal indirect support*

Universal supportedness
$(\forall x)(\exists y)(ySx)$, or equivalently $(\forall x)(\exists y)(yS^{*}x)$

Universal supportedness requires that E has no unsupported element This seems to be a necessary condition for coherentism However, as can be seen from Figure 1, it is not a sufficient condition The structure shown in that figure satisfies *Universal supportedness*, but it is a

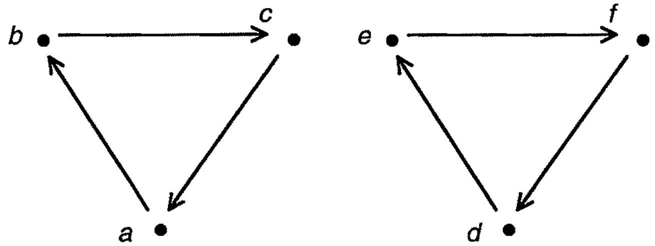


Figure 2

clear example of a structure that we would call foundationalist rather than coherentist.

Universal supportedness can be combined with the following condition, that is not satisfied by the pattern shown in Figure 1.

Universal supportingness
$(\forall x)(\exists y)(xSy)$, or equivalently $(\forall x)(\exists y)(xS^{*}y)$

Both *Universal supportedness* and *Universal supportingness* follow from *Universal indirect support*, and they are both important aspects of what we usually mean by coherentism. However, even in combination they are not sufficient to define a reasonable notion of coherence. This can be seen from Figure 2. What is missing here is a property of a coherent set that has been emphasized by several authors on the subject. There should be no isolated subsystem, or part that is unconnected with the rest of the system.[5] We can call this a requirement of nonfragmentation. Let $E=E_1 \cup E_2$, where $E_1 \cap E_2 = \varnothing$. If there is no support from (any element of) E_1 to (any element of) E_2, and no support from E_2 to E_1, then they are two separate subsystems, and E is fragmented. For this *not* to be the case, support in one of the directions is sufficient. If all elements of E_2 are supported by several elements of E_1, but there is no support in the opposite direction, then it would be wrong to say that $\langle E,S\rangle$ is fragmented or that it has isolated subsystems.

The following is a simple condition that approximates the requirement that E contains no isolated subsystem.[6]

Nonfragmentation
$(\exists x)(\forall y)(xS^{*}y)$

---

[5] Laurence BonJour, *The Structure of Empirical Knowledge* (Cambridge: Harvard, 1985), p. 97. Wolfgang Spohn, "Two Coherence Principles," *Erkenntnis* L, 2–3 (1999): 155–75.

[6] Arguably, it is simple at the price of being too strong. An interesting weaker alternative is $(\forall x)(\forall y)(x\overset{\leftrightarrow}{S}{}^{*}y)$, or equivalently $(\exists x)(\forall y)(x\overset{\leftrightarrow}{S}{}^{*}y)$, where $x\overset{\leftrightarrow}{S}y$ holds if and only if either xSy or ySx.

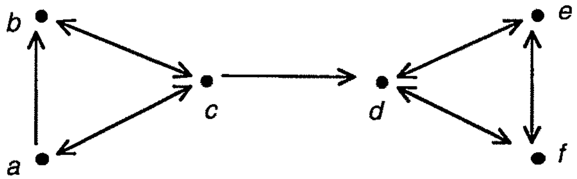


Figure 3

Figure 3 shows a case in which *Universal supportedness*, *Universal supportingness*, and *Nonfragmentation* are all satisfied, whereas *Universal indirect support* is not The combination of these three conditions ensures coherence to a certain degree, but not to the (arguably unrealistically) high degree ensured by *Universal indirect support* Therefore, we can distinguish between two coherence requirements The strong coherence requirement coincides with *Universal indirect support* The weak coherence requirement consists of the three conditions *Universal supportedness*, *Universal supportingness*, and *Nonfragmentation*

III WHAT IS FOUNDATIONALISM?

A foundationalist structure should be one in which a certain subset of *E*, the base, can be identified that has specific properties in relation to the rest of the set Which are these properties? Sosa's formulation of the pyramid metaphor, quoted in section I, provides us with two such properties The base is a part "whose stability and security does not derive from the upper stories or sections," in other words the base *is not supported by* the upper layers Furthermore it *supports* these other parts, thereby forming their "stable and secure foundations "

The first of these two properties will be satisfied if we require that the elements of the base are not supported by any other elements of *E*[7] It would then follow that *E* can have a base only if it satisfies the following condition

Unsupported elements
$(\exists x)(\forall y)\neg(ySx)$ or equivalently $(\exists x)(\forall y)\neg(yS^{*}x)$

Unsupported elements is well in line with traditional foundationalism Nevertheless, it is in fact quite implausible Consider Figure 4 Here $\{a, b\}$ can be said to be a base for *E*, although the two elements of this base support each other

---

[7] They are typically assumed to have some sort of external support, that is, support from nonelements of *E*, but such external support is not represented by *S*

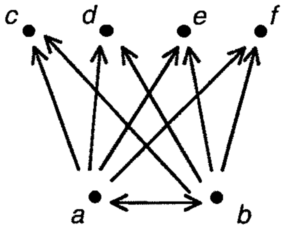


Figure 4

Foundationalism is closely related to the view that some sort of first principles are needed to avoid the two evils circularity (that is not considered an evil by coherentists) and infinite regress. To take a very simple example, our sense experiences may be taken as the basis of our empirically based knowledge. This means that these sense experiences cannot in their turn be based on some further type of source or justification. However, it does not follow that our sense experiences cannot support each other. To the contrary, since our senses provide us only with incomplete and fallible information, combinations of sense impressions are often essential. Hence, when you play a musical instrument, the auditory, tactile, and visual feedback that you receive from your actions on the instrument combine to provide you with the information you need on how to proceed.[8] More generally speaking, connections of support among the elements of a base strengthen the base, so that it provides more secure foundations for that which is built upon it. The idea that first principles should be independent of each other's support seems to be a remnant of philosophical speculations on sense data that have since long been refuted by science. In actual life, reliable empirical data typically support each other. Arguably, so do tenable moral principles. Therefore, *Unsupported elements* should not in general be required in a foundationalist structure. It can however be replaced by the much more reasonable requirement that the base is not supported by any of the elements of E outside of the base, hence

Unsupported set
There is a set E' with $\emptyset \neq E' \subset E$, such that
$(\forall y \in E \setminus E')(\forall x \in E')\neg(ySx)$, or equivalently $(\forall y \in E \setminus E')(\forall x \in E')\neg(yS^*x)$

---

[8] Bengt Edlund, "Structural Symmetry and Proprioceptive Patterns in Music," *Symmetry Culture and Science*, VII, 2 (1996) 139–51

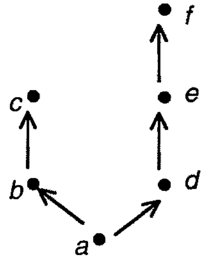


Figure 5

This condition is clearly not sufficient to characterize a foundationalist structure Just like *Unsupported elements*, it is satisfied if S is empty We also need a positive requirement on the base E', guaranteeing that it supports the rest of E This is the other of Sosa's two conditions on the base of the pyramid It can be added as an extension of *Unsupported set*

Base
There is a set E' with $\emptyset \neq E' \subset E$, such that
$(\forall y \in E \setminus E')(\forall x \in E')\neg(yS^*x)$ and
$(\forall y \in E \setminus E')(\exists x \in E')(xS^*y)$

This is a reasonable definition of a base, saying essentially that it supports the rest of the structure but is not supported by it It should be noted, however, that with this definition, the base will not in all cases be uniquely defined Hence, in Figure 5, there are eleven alternative bases according to *Base*[9] A reasonable way to avoid this multiplicity of bases is to settle for the inclusion-minimal base The following observation shows that there is always a unique inclusion-minimal base

OBSERVATION 3 Let each of E_1 and E_2 be a minimal base of E, that is, each of them satisfies the criteria of *Base* but has no proper subset that does so Then $E_1 = E_2$

Base can be strengthened so that it implies the above-mentioned requirement *Unsupported elements*

Disconnected base
There is a set E' with $\emptyset \neq E' \subset E$, such that
$(\forall y)(\forall x \in E')\neg(yS^*x)$ and
$(\forall y \in E \setminus E')(\exists x \in E')(xS^*y)$

---

[9] $\{a\}$, $\{a,d\}$, $\{a,d,e\}$, $\{a,d,e,f\}$, $\{a,b\}$, $\{a,b,d\}$, $\{a,b,d,e\}$, $\{a,b,d,e,f\}$, $\{a,b,c\}$, $\{a,b,c,d\}$, and $\{a,b,c,d,e,f\}$

CENTRAL LIBRARY
UNIVERSITY OF CALCUTTA
CAMPUS ORE

We can distinguish between a weak and a strong criterion for foundationalism, namely *Base* respectively *Disconnected base* The strong criterion conforms with foundationalist tradition in requiring that the elements of the base do not support each other, but as we have seen this is not a very well-motivated requirement

IV CONCLUSION

Summarizing the above, we have weak and strong requirements for both coherentism and foundationalism

Weak coherence = *Universal supportedness, Universal supportingness,* and *Nonfragmentation*

Strong coherence = *Universal indirect support*

Weak foundationalism = *Base*

Strong foundationalism = *Disconnected base*

OBSERVATION 4 (1) Weak coherence and weak foundationalism are compatible
(2) Strong coherence and weak foundationalism are incompatible
(3) Weak coherence and strong foundationalism are incompatible

The compatibility of weak coherence and weak foundationalism opens up the possibility of considering structures of support that are intermediate between, or combine, the classical notions of coherentism and foundationalism Arguably, such intermediate structures may in some contexts be more plausible than the uncompromisingly foundationalist or coherentist ones As we saw in sections II and III, neither strong foundationalism nor strong coherentism has any clear advantage over the weak variant in terms of intuitive plausibility

Therefore, the traditional dichotomy between coherentism and foundationalism is misleading Instead of classifying support structures according to this dichotomy, the many variations and nuances in actual structures of support should be investigated Realistic support structures may have a combination of coherentist and foundationalist properties In particular, they may satisfy both the weak foundationalism condition (*Base*) and the weak coherence condition (*Universal supportedness, Universal supportingness,* and *Nonfragmentation*) The support pattern illustrated in Figure 6 exemplifies this combination

For one example of this, let $\langle E,S \rangle$ represent the system consisting of our beliefs and their justificatory interrelations There are good reasons to represent our set of beliefs as a logically closed set of sentences However, as a consequence of logical closure, that set will

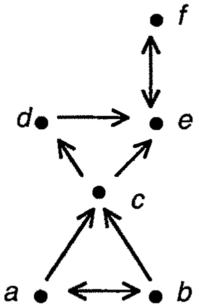


Figure 6

contain some belief-representing sentences that have no other justification than that of being logical consequences of other beliefs that have more weighty justifications Hence, I recently acquired the belief that my bicycle has a flat tire (p) As a consequence of this I am also committed to the belief that either my bicycle has a flat tire or I am the Emperor of Japan ($p \vee q$) Beliefs of the latter type do not contribute to the justification of beliefs of the former type The belief system can plausibly be modelled as satisfying both weak coherentism and weak foundationalism, with the base consisting of beliefs of the former type, that is, beliefs that are not merely derived This is admittedly a quite different type of base than what foundationalist epistemologists have had in mind However, since weak foundationalism allows for several bases, this leaves the question open whether there is also a more restricted base of traditional nature in the belief system

SVEN OVE HANSSON

Royal Institute of Technology

APPENDIX PROOFS

PROOF OF OBSERVATION 1 *Part 1* Suppose to the contrary that x, $v \in E \setminus \mathrm{Cn}(\emptyset)$ and that v is equivalent with $x \rightarrow y$ Then due to *Universal direct support*, xSv, and due to *Deductive relevance* there is some z such that

$z \nvdash v$ and $x \& z \vdash v$
$z \nvdash x \rightarrow y$ and $x \& z \vdash x \rightarrow y$
$z \nvdash x \rightarrow y$ and $z \vdash x \rightarrow (x \rightarrow y)$

which is impossible since $x \rightarrow y$ and $x \rightarrow (x \rightarrow y)$ are equivalent

Part 2 Suppose to the contrary that *Universal direct support* and *Deductive relevance* are satisfied Since E contains at least three nonequivalent sentences it contains two nontautologies x and y such that $x \rightarrow y$ is not a tautology It follows from part 1 that $x \rightarrow y \notin E \backslash \mathrm{Cn}(\varnothing)$, contrary to the logical closure of E

PROOF OF OBSERVATION 2 There are three cases *Case 1,* $\vdash x$ or $\vdash y$ Then xSy and thus xS^*y

Case 2, $\nvdash x$, $\nvdash y$, and $\vdash x \vee y$ Then since $\nvdash x$ we have $\nvdash y \rightarrow x$, and similarly $\nvdash y$ yields $\nvdash x \rightarrow y$ Thus

$xS(x\&y)$ since $y \nvdash x\&y$ and $x\&y \vdash x\&y$
$(x\&y)Sy$ since $x \nvdash y$ and $x\&(x\&y) \vdash y$
It follows from $xS(x\&y)$ and $(x\&y)Sy$ that xS^*y

Case 3, $\nvdash x \vee y$ Then equivalently $\nvdash (x \rightarrow y) \rightarrow y$ We have xSy since $x \rightarrow y \nvdash y$ and $x\&(x \rightarrow y) \vdash y$

PROOF OF OBSERVATION 3 Suppose to the contrary that $E_1 \neq E_2$ Due to the minimality condition, both $E_1 \backslash E_2$ and $E_2 \backslash E_1$ are nonempty

Now suppose that for all $y \in E_2 \backslash E_1$ there is some $x \in E_1 \cap E_2$ such that xS^*y Then (1) since E_2 is a base, for all $y \in E \backslash (E_1 \cap E_2)$ there is some $x \in E_1 \cap E_2$ such that xS^*y Furthermore, (2) since E_2 is a base, for all $y \in E \backslash E_1 \cap E_2$ and $x \in E_1 \cap E_2$ we have $\neg(zS^*x)$, and (3) since E_1 is a base, for all $z \in E \backslash E_1$ and $x \in E_1 \cap E_2$ we have $\neg(zS^*x)$ It follows from (2) and (3) that for all $z \in E \backslash (E_1 \cap E_2)$ and $x \in E_1 \cap E_2$ we have $\neg(zS^*x)$ From this and (1) it follows that $E_1 \cap E_2$ is a base, contrary to the conditions

Hence, the assumption was wrong, and consequently there is some $y \in E_2 \backslash E_1$ such that $\neg(xS^*y)$ for all $x \in E_1 \cap E_2$ Since E_1 is a base there must be some $x \in E_1$ such that xS^*y, thus $x \in E_1 \backslash (E_1 \cap E_2)$, that is, $x \in E_1 \backslash E_2$ But then E_2 is supported by an element of $E \backslash E_2$, so that E_2 is not a base, contrary to the conditions Contradiction

PROOF OF OBSERVATION 4 *Part 1* See Figure 3 or 6

Part 2 *Universal indirect support* contradicts the first condition of *Base*

Part 3 The first condition of *Disconnected base* contradicts *Universal supportedness*

LOVE AND DEATH

The Ecclesiast, purveyor of dyspeptic wisdom, reflects on the misfortunes of the dead "The dead know nothing, they have no more reward and even the memory of them is lost" (9 5) If at least the memory lingered on—that would be something But the Ecclesiast uncharacteristically misses an opportunity to rub it in, for it can actually seem much worse it is not just that we are forgotten in the vast eternities of time, it is that even those we love[1] and who love us the most display an astonishing facility for "moving on," for adapting to their loss and so avoiding significant distress over our demise Moreover, wives and husbands remarry, often at depressingly brief intervals, further diminishing the significance of the loss to them, we retain photographs and other trinkets of remembrance, but all of the emotions that were once insurmountable barriers to similar relationships with others are irrecoverably lost All of this occurs because of adaptive processes that may be in the self-interest of the survivor since they facilitate a rapid return to a healthy emotional baseline, but they involve profound reasons for regret as well, both because of what these processes say about the bereaved and, more generally, what they say about the significance of our relationships with one another At least, that is one interpretation of love and death I find this gloomy view both compelling and depressing, and it is worth expounding it in more detail for the light it sheds on our lives and our deaths

I

How *do* spouses or partners in a long-term, committed relationship react when the person they love dies? In answering this question, we may be tempted to form an estimate by focusing on what we imagine we might feel immediately after hearing the news that our husband or wife had died But this and other armchair estimates of emotional responses are notoriously unreliable, and fortunately there exists a large empirical literature on the psychology of bereavement that can serve as a fixed point The most important finding of this body of work is summarized by a leading researcher

> [M]any, and sometimes the majority, of bereaved individuals exhibit only short-lived grief-reactions and a relatively rapid return to baseline

---

[1] The focus here will be on "romantic" love, though much of what I say applies to other forms as well

0022-362X/07/0406/301–16

> functioning Bereavement theorists have tended to assume that the normative responses to loss involve either chronic suffering or gradual recovery lasting at least several years, the relative absence of distress during bereavement is thought be both rare and psychopathological [But] recent research has provided a strong challenge to these views The relative absence of grief symptoms and the continued ability to function adequately following the death of a close relation do not appear to reflect denial or pathology but rather an inherent and adaptive resilience in the face of loss [2]

The results of empirical investigation thus seem to conflict with a widely held view in our culture that the loss of a partner or spouse is invariably or at least usually an agonizing blow with long-lasting and significant impact Contrary to this folk view (and certain nonempirical bereavement theories), empirical research seems to show that most people manifest what the author above refers to as *resilience* in the face of loss although they are initially traumatized, they quickly recover and manifest little long-term distress And, again contrary to folk wisdom, this does not seem to be the result of repression or of having had an unfulfilling relationship, most people simply adapt far better to their loss than we tend to believe

Nor is this finding at all unusual, indeed, it represents the consensus of nearly all of the empirical research dedicated to testing how people actually react to the loss of a spouse The author cites another study in which nearly half of the participants did not show even mild depression following a spousal loss [3] Another researcher who carried out a long-term investigation reports that, compared to a control group, "The effect of bereavement on symptoms of depression and general psychopathology was significant only at 2 months following loss "[4] Still another expert sums up his work, "A general conclusion of this study is that the death of a spouse in later life does impact the surviving spouse's subjective well-being but not to the extent that many would expect "[5] If these results are surprising, it gets stranger still research indicates that not only do half or more of bereaved spouses tend to be resilient or muted in their reaction to their loss, but a consistent 10% or so of the bereaved experience a dramatic *increase* in subjective well-being following the loss.[6] And prospective studies that allow the tracking of well-being before as well as after a loss reveal that a large proportion of those who *do* report being deeply unhappy several months after their loss were deeply unhappy before the loss as well; in other words, it is unclear that their distress is caused by their loss rather than standing features of personality or circumstance.[7] Moreover, with some minor differences these results appear to apply cross-culturally.[8] Finally, given this muted reaction, it is no surprise that the bereaved often find it fairly easy to remarry within a fairly short period of time. Although there are somewhat different figures given in the literature, owing perhaps to different investigative methods, it is clear that those who are young and therefore find it relatively easy to secure new partners remarry quite soon after their loss.[9] And of course we must bear in mind that even among the elderly who remarry far less often, the *desire* to remarry or date is much higher than the actual rate, both for generic reasons and because of an increasing gender imbalance with age increase.[10]

---

[2] G Bonanno et al, "Resilience to Loss in Bereaved Spouses, Bereaved Parents, and Bereaved Gay Men," *The Journal of Personality and Social Psychology*, LXXXVIII (2005) 827–43, quoted from p 827, citations omitted

[3] The study is S Zisook et al, "The Many Faces of Depression Following Spousal Bereavement," *Journal of Affective Disorders*, XLV (1997) 85–94

[4] A Futterman, J Peterson, and M Gilewski, "The Effects of Late-Life Spousal Bereavement over a 30-Month Interval," *Psychology and Aging*, VI (1991) 434–41, on p 438

[5] Dale Lund, "Impact of Spousal Bereavement on the Subjective Well-Being of Older Adults," in Lund et al, eds, *Older Bereaved Spouses* (New York Hemisphere, 1989), pp 3–15, on p 12, cited in Jeremy Blumenthal, "Law and the Emotions The Problems of Affective Forecasting," *Indiana Law Journal*, LXXX (2004) 1–95

It is important to concede the other side of the story and not to exaggerate these findings, summarized in extreme brevity just now. Obviously a large number of people in absolute terms are completely devastated by their loss, and some never recover. Loss of a spouse is in fact associated with a sharp rise in risk of death, especially shortly after the loss, indicating that the consequences are anything but trivial.[11]

---

[6] Bonanno, R. Nesse, and C. Wortman, "Prospective Patterns of Resilience and Maladjustment during Widowhood," *Psychology and Aging*, XIX (2004): 260–71.

[7] Bonanno et al., "Resilience to Loss and Chronic Grief: A Prospective Study from Preloss to 18 Months," *The Journal of Personality and Social Psychology*, LXXXIII (2002): 1150–64.

[8] Bonanno et al., "Grief Processing and Deliberate Grief Avoidance: A Prospective Comparison of Bereaved Spouses and Parents in the United States and the People's Republic of China," *Journal of Consulting and Clinical Psychology*, LXXIII (2005): 86–98.

[9] The median interval for remarriage among whites is 1.7 years for men and 3.5 for women. About 90% of white widows under 45 remarry, with a similar pattern holding for women as well until the mid 30s, at which point there is a steep drop-off in probability of remarriage. Interestingly, these figures are respectively higher and lower among blacks. See W. Cleveland and D. Gianturco, "Remarriage Probability after Widowhood," *Journal of Gerontology*, XXXI (1976): 99–103. Naturally, remarriage looks very different among the aged. For a recent study, see K. Smith, C. Zick, and G. Duncan, "Remarriage Patterns among Recent Widows and Widowers," *Demographics*, XXVIII (1991): 361–74.

[10] D. Carr, "The Desire to Date and Remarry among Older Widows and Widowers," *The Journal of Marriage and Family*, LXVI (2004): 1051–68.

[11] N. Christakis and P. Allison, "Mortality after the Hospitalization of a Spouse," *New England Journal of Medicine*, CCCLIV (2006): 719–30. It is important to note that the

Even when subjects report relatively little psychic trauma, they do sometimes report grieving for a substantial length of time, though this reported grieving appears to have little connection with measures of happiness or subjective well-being (and thus does not correspond, perhaps, to grief in the way it is usually understood) [12] Nevertheless, the cumulative evidence seems to show quite clearly that most people do not experience significant long-term distress when they lose the person they have committed their lives to It is important to note, also, that the investigations cited for the most part rely on self-reports of grief, either in response to in-person interviews or more usually responses to paper-and-pencil surveys The direction of exaggeration is thus very likely to be in the direction of overreporting grief and distress since this is what we culturally expect of those experiencing such a loss In fact, this expectation is so strong that increased grief display is positively correlated with likeability ratings [13]

Like all empirical findings, these results are subject to revision and refutation I will assume that they are roughly accurate, but even if they turn out not to be, if we like we can simply stipulate that the discussion applies to the (majority) group whose reaction to the loss ranges from muted or resilient to ebullient This provides a certain immunity to future revision, since presumably only the size but not the existence of this group is likely to be challenged In any case, later on I will try and show that these empirical findings, while useful for motivating the discussion, are not essential to my claims

II

Some may find these results encouraging and only encouraging a large class of people whose lives we thought were shattered by a major loss turn out to be affected only to a fairly low degree There is less misery in the world than we thought If this response is unadulterated by qualms or hedges, it may be rooted in a general view about how to evaluate reactions to losses the best reaction to a loss is the one that most promotes the interests of the one suffering the loss Call this the *Adaptive Theory* This theory may in turn find some support in a still

---

subjects of this study were quite old—the average age was over 70—and also that the same effect was observed in nonlethal hospitalizations For example, the spousal risk to men whose wives have been hospitalized with a hip injury is comparable to the risk associated with a wife's death, especially after the first few months This indicates that what is at issue is probably not extreme grief or depression but more mundane problems arising from interdependence

[12] Bonanno et al , "Resilience to Loss in Bereaved Spouses "

[13] Bonanno et al , "Resilience to Loss in Bereaved Spouses "

more general conception of human behavior—one that says that the way to evaluate the rationality of *any* human activity is in terms of its propensity to promote the person's interests (The self-interest theory of rationality is one manifestation of this normative attitude toward behavior [14]) Since many people suppose that unpleasant experiences of grief and psychological distress are contrary to our interests, they might infer from the Adaptive Theory that the response most people actually seem to have is close to the best response, in failing to be significantly distressed, the bereaved are coping in a near-maximally adaptive fashion Since the evidence indicates that there is no penalty to be paid for failing to suffer from a loss, the best response to grief may thus seem to be little or none at all

However, many people are not only surprised but shocked and dismayed by these results, as the tendency toward greater dislike for those who display comparatively little grief brings out Why is that? What is it that might make us uncomfortable about resilience? One reason we should consider is that insufficient grief following a loss may indicate that the one suffering the loss never valued the object sufficiently in the first place We may imagine dying and infer from the subjunctive truth that those we love would probably not be deeply affected by our absence that we are not valued very highly even now To the extent that responses to losses and valuing of the thing lost go together, it can seem tempting to move from disappointment with the one to suspicion about the other If this were right, we would need to think harder about the use of "best" in the Adaptive Theory Perhaps resilience is best for promoting the person's own self-interest, but that may only show that in this domain, as in so many others, there are considerations counting against having the attitudes that best promote our interests

There is, however, a powerful objection to this line of thought For what is true of reactions to the loss of a spouse turns out to be true of our reactions to other major losses as well We are systematically mistaken in what is sometime called our "affective forecasting" [15] what we imagine to be catastrophes for us routinely turn out to be only minor blips that we quickly overcome Being denied tenure, for instance, has very little effect on long-term happiness [16] Contrary to what most academics imagine, whether or not they are denied tenure is unlikely to make much of a difference in their subjective well-being after an initial and short-lived dip. Similarly, dissolving a romantic relationship, losing a child,[17] sustaining a debilitating spinal-chord injury,[18] or other major medical problems[19] may produce differences compared to control groups, but these are often surprisingly small and tend to diminish or even disappear after a short while. (Conversely, we are also wrong about things like gaining tenure or winning the lottery, which tend not to produce as much happiness as we predict.) The reasons for these rapid recoveries are not fully understood and seem to be heterogeneous.[20] However, some of the mechanisms at work are fairly obvious, particularly various contrast effects. Recent improvements or declines in well-being contrast with still-vivid memories of previous states and affect us strongly, but after a short while this contrast effect wears off and we get used to what we now regard as the new baseline. Also, sudden changes can reduce the significance of previously important-seeming events, so that, say, simple pleasures like going to the movies suddenly seem less exciting once we have won the lottery. For these and other sorts of reasons, it appears to be difficult permanently to raise or lower people's levels of happiness provided there are not continual "shocks," and provided a certain threshold has been reached. This view is supported by the finding that, when it comes to events (as opposed to persistent states, like being clinically depressed), almost nothing that happens to us has a significant impact on happiness beyond three months or so.[21]

---

[14] See Derek Parfit's *Reasons and Persons* (New York Oxford, 1984) for an extended discussion and critique of the self-interest theory

[15] See the account and further references given in D Gilbert and J Ebert, "Decisions and Revisions The Affective Forecasting of Changeable Outcomes," *Journal of Personality and Social Psychology*, LXXXII (2002) 503–14

[16] D Gilbert et al , "Immune Neglect A Source of Durability Bias in Affective Forecasting," *The Journal of Personality and Social Psychology*, LXXV (1998) 617–38

What this shows is that even when it comes to paradigm cases of things we care very deeply about (for example, our spinal chords) our reaction to loss is muted and far less than we tend to imagine it would be. This seems to suggest that we cannot explain dissatisfaction with resilience to the loss of a spouse by claiming that resilience indicates apathy. Moreover, these findings imply that any plausible attack on the Adaptive Theory must account for the fact that our reactions to bereavement form part of a general pattern of resilience toward even extreme and disastrous changes The Adaptive Theorist can now point out that resilience to losing those we love is part of a disposition that plays a deep and systematic role in making us the kinds of creatures that can overcome the frequent and inevitable setbacks that we must suffer over a lifetime

---

[17] Bonanno et al., "Resilience to Loss in Bereaved Spouses."

[18] P. Brickman, D. Coates, and R. Janoff-Bulman, "Lottery Winners and Accident Victims: Is Happiness Relative?" *The Journal of Personality and Social Psychology*, xxxvi (1978): 917–27.

[19] For an important study and further references, see J. Riis et al., "Ignorance of Hedonic Adaptation to Hemodialysis," *Journal of Experimental Psychology: General*, cxxxiv, 1 (2005): 3–9.

[20] Bonanno, "Resilience in the Face of Potential Trauma," *Current Directions in Psychological Science*, xiv, 3 (2005): 135–38.

[21] E. Suh, E. Diener, and F. Fujita, "Events and Subjective Well-Being: Only Recent Events Matter," *The Journal of Personality and Social Psychology*, lxx (1996): 1091–102.

Of course, the point about our not caring enough about what we lose could be broadened to take this larger pattern into account An embittered cynic might say, "I used to think that reactions to the loss of a spouse were especially objectionable These new facts show that not to be the case But now things are worse Apparently we do not care very much about *anything* The facts about our reactions to losses show that evolution has endowed us with a deplorably superficial capacity for valuing things in general My sorrow at learning that lovers don't care enough about each other to be significantly distressed by the loss of the other isn't mitigated by learning that they don't care that much about their spinal chords either " This extreme position might be buttressed by claiming that part of what it *is* to care deeply about something is to be disposed to suffer deeply following losses "A person who cares about something is, as it were, invested in it He *identifies* himself with what he cares about in the sense that he makes himself vulnerable to losses "[22] If we prove *in*vulnerable or insufficiently vulnerable, that may ground suspicion that we did not deeply care to begin with To put it another way, facts about caring have conditional implications involving responses to losses, and we can reason from the failure of these conditionals to be satisfied to the absence of concern This may be reflected, for instance, in the difficulty we have in accepting statements like this "You are the love and light of my life, but naturally, were you to die, it would have little significant impact on me beyond a few months, and I would probably fall in love with someone else shortly thereafter "[23]

Apart from being implausibly extreme in implying that none of us cares about much of anything, this view overlooks the massive evidence that we do care deeply about people before they die This concern is manifested in our willingness to make sacrifices for them, our pain in the face of even their minor misfortunes, and our general disposition to protect and advance the interests of those we love The extreme view also overlooks different reasons for failing to be devastated by loss, only one of which is failing to care about something to begin with. Take denial of tenure. The reason people are so wrong about their predicted response to this event is not that they care less about tenure than they think, but rather that they overlook the elaborate psychological apparatus we possess for dealing with such events.[24] After their initial horror, unfortunate junior professors soon begin to rationalize their failure by appealing to baroque conspiracy theories, departmental politics, and so on (never the inadequacy of their work), and within a few months the contrast between a lousy present and better immediate past is lost in any case. Given the prima facie evidence in favor of the antecedent high concern for tenure, the best explanation for lack of distress over denial of tenure is thus likely to be the engaging of various adaptive mechanisms, not a prior absence of concern. Although the bereaved are different to the extent that rationalization and other coping mechanisms of junior professors probably do not apply to them, they too can say that their lack of distress merely indicates that concern for a person or object can be severed from distress over its loss by adaptive mechanisms. It is these mechanisms that explain how someone could be willing to risk her life for her husband while failing to be significantly traumatized by his death. It just turns out to be a remarkable trait of our species that caring very deeply about someone is compatible with a strongly muted reaction to their death. Friends of the Adaptive Theory can thus not only claim that our response to loss is best, but that so far no ground for genuine ill-ease with that response has been located.

---

[22] Harry Frankfurt, "The Importance of What We Care About," in *The Importance of What We Care About* (New York Cambridge, 1988), pp 80–94, on p 83

[23] Since I go on to rebut this view on other grounds, I will not tarry over the possibility that the inconsistency involves mere conversational pragmatics

The extreme response is a dead end; resilience is evidence of a post-factum ability to adapt to loss, not an antecedent lack of concern. Nonetheless, there is something right about the extreme response. Resilience does tell us something important and distressing about our relationships with those we love, even if the precise feature at issue has been misidentified by the extremist. One source of anxiety concerning our adaptive capabilities lies, I believe, in what they imply about our *importance* for those we love. The problem is that because our deaths make a comparatively minor impact on their lives, we may feel forced to conclude that we do not possess the kind of importance for them that we thought we had. The notion of "importance" that is relevant here, or what we might in some contexts call "significance" or "value" notwithstanding subtle differences between the three, is a hard one to pin down, but a few simple examples will make clear what is at issue. Our importance to an organization like a baseball team or Congress is great when we make an enormous difference to its operations, when our absence wreaks havoc, and when we are unique and irreplaceable in what we do. Conversely, claims of importance or significance are inflated when it turns out that nothing we do really matters or that a year's leave of absence would go unnoticed and we could be easily replaced. Similarly, in the personal sphere we need and therefore value relationships with some people more than others. Friends who add greatly to our lives seem important to us, while the pizza delivery boy is insignificant to us since he does not do anything we need all that much and even the little he does is easily done by someone else. (As this example shows, the insignificance of something *for us*—relative to certain aims and preferences—need not imply anything about the value of the object as such. Nothing I say here is incompatible with the infinite Kantian worth of rational pizza boys.)

---

[24] That is, we neglect our "affective immune system." See Gilbert et al., "Immune Neglect."

An important feature of *importance* is thus that the concept has a counterfactual dimension, my supposed importance to a baseball team can be undercut by what happens after I leave it. It is for this reason that worries about our significance in the face of resilience cannot be quelled by the point made earlier against the extremist: evidence that we care for our lovers while they live is not undermined by facts about what happens after they die, but the same is not true of importance. A muted response to our death *is* relevant to our importance while we still live, since counterfactuals in part determine the importance or value of something. We like to believe that we are *needed* by our husband or wife and that consequently losing us should have a profound and lasting effect on them, just as the sudden injury of a key baseball player should have a disruptive and debilitating effect on the team. Most of us tend to assume that our relationship with the ones we love is so important to them that severing that relationship would make a deep impact on their ability to continue to lead happy worthwhile lives. The fact that our beliefs about these matters are false and that our loved ones are resilient to the loss of us seems to show that we do not have the significance that we thought we did.

We can explain what is distressing about the high probability of and brief interval before remarriage along similar lines. Not being *needed* is one way to fail to be important, but of course not the only way, since even people performing essential functions can be *replaced*, and fungibility is antithetical to importance or value. (That is part of the reason for the well-known tendency of leaders such as founders of businesses or Roman emperors to fail to institute succession plans: the existence of such a plan immediately reduces the leader's own importance.) Of course, it would be not only crass but inaccurate to think of remarriage as replacing the deceased *in toto*. Both people and the specific contours of their relationships are unique and will never be wholly mimicked by successor-relationships. But the same is true of baseball players and Congressmen: a team will function somewhat differently once it loses a wily southpaw, and Congress will never be quite the same without Preston Brooks and his cane. The reason the operations of these organizations are not substantially impeded by such losses is that the differences that matter to organizations are determined by their overall goals and functioning (winning ball games and legislating effectively), and these are not much affected by mere idiosyncrasies. Contrary to the suggestion made earlier, at least some spouses and lovers may need us very much. But the tendency toward remarriage combined with the comparatively low rates of distress among the bereaved suggest that the ways in which we are needed are functional rather than idiosyncratic, we play an important role in the lives of those we love, but not a role that others cannot with relative ease fill once we are gone, since our unique traits do make much of a difference to our success in those roles. We may be desperately needed as companions, friends, sex partners and intimates, but these roles endow us with much less significance than we imagine, given that we can be functionally replaced in these respects, and—so the evidence indicates—upon our deaths very often are. This is not to say that those who remarry do not have a special place in their heart for their previous spouse that will never be touched by his or her successor, it is only to explain why the adaptive responses of those we love can give us reason for regret, since they seem to show that we do not have the importance for our husbands and wives that we thought.

But it is not only those contemplating their own demise and the lack of impact that event will have on those they love who have reason to be uneasy about resilience, for despite the obvious advantages the Adaptive Theory draws our attention to, even those who have *lost* their spouse, and so directly benefit from adaptation, may have reason to regret the benefits resilience confers. Our "emotional immune system," while promoting our interests by allowing us to continue functioning in the face of trauma, also renders us unable to take in and register fully the significance of our losses. Part of what being the vulnerable creatures of flesh and blood that we are means is that we are subject to staggering losses in the form of the deaths of those we love, and yet our reaction to those losses is utterly incommensurate with their value, especially after the first month or two have passed. The good of a happy relationship with a lover is one that we value more highly than almost any other, and yet when we lose that good, our response over time does not seem to reflect its preciousness to us

Resilience thus seems to deprive us of our ability to care about those we love to their full measure after they are gone, and so deprives us of insight into our own condition.

This claim may seem peculiar. Assuming we are not in some pathological state of denial and are thus in possession of the appropriate beliefs, how can we possibly fail to register our losses? The answer to this lies in the difference between the superficial discrimination of some fact and what we sometimes describe by saying that someone has really "absorbed" or "taken in" something—a difference that typically involves having certain emotions in connection with what has been registered. Explaining what exactly the relationship between having emotional responses is to digesting fully various goods and bads that befall us is admittedly difficult, since on reflection it is far from clear why there should be an affective component to processing certain experiences. Why need feelings accompany the purely cognitive processes of discrimination in order for us to believe that someone has really "gotten it" when, say, she gets word of a death or an amputation or a (welcome) marriage proposal? Perhaps part of it is that emotions sometimes seem to have a role analogous to sense organs for us: the emotions are part of the means by which we perceive value.[25] Of course, emotions also do other things, and cognitive processes like practical reasoning can play a role in perceiving value as well, but it is nonetheless hard to envision someone being fully capable of recognizing a good or a bad without the appropriate emotional responses. In fact, there is powerful empirical evidence to suggest that without emotional responses we find it so difficult to recognize value that we cannot make even elementary decisions that require us to see one option as better than another.[26] Similarly, psychopaths seem unable to grasp the moral significance of their victims despite possessing normal cognitive abilities, and the best evidence is that this is due to their inability to experience certain emotions.[27] Those with affective disorders thus seem unable to grasp whether certain things really *matter*. If these remarks are on track, then our adaptive but muted responses prevent us from forming a correct picture of how things really are, value-wise. In a sense, resilience is a form of benign—or at least adaptive—blindness.[28]

---

[25] Some will want to follow Hume in saying that the emotions are the means by which we *project* value (for example, second *Enquiry*, Appendix I). But Humeans can and should lament the inability to project in roughly the same way anti-Humeans lament the inability to perceive value; it is not as if Humeans are less appalled by those afflicted with affective disorders which prevent them from reacting to important changes in well-being. Their explanation will simply take a different form. For discussion, see Simon Blackburn's *Ruling Passions* (New York: Oxford, 1998), especially chapter 9. For a defense of the view that emotions can represent things like losses and that emotions involve perceptions of our relation to the world, see Jesse Prinz, *Gut Reactions: A Perceptual Theory of Emotion* (New York: Oxford, 2004), chapters 3 (especially pp. 60–66) and 10. Prinz also defends the assumption I make here that emotions involve feelings and are not purely cognitive.

[26] For instance, subjects with affective disorders often fail at certain games of chance because they do not see bad outcomes as bad anymore. See Antonio Damasio, *Descartes' Error* (New York: Avon, 1994), especially chapters 3 and 9.

But even if this were right, it may not seem obvious that such a blindness or inability fully to comprehend the loss of a good is a bad thing. The loss itself is one bad, and experiencing intense psychological distress in response to it simply adds a second; how can suffering one but not the second bad be *worse* than suffering both? The oddity comes out especially well in Proust's statement of the view I am defending:

> Our dread of a future in which we must forego the sight of faces and the sound of voices which we love and from which today we derive our dearest joy, this dread, far from being dissipated, is intensified, if to the pain of such a privation we feel that there will be added what seems to us now in anticipation more painful still: not to feel it as a pain at all—to remain indifferent.[29]

We can explicate the feeling Proust refers to by noting that failing to recognize how great a good we have irretrievably lost means that we are to some extent deluded about our own condition. That seems undesirable for the same reason other forms of pleasant delusion are often undesirable. To the extent that we care about being aware of our general condition, we have grounds to regret even adaptive syndromes which impede that awareness. (Notice that such a lack of emotional awareness can seem equally disturbing when it involves something good, as in an autobiography that refers to "unaccountably streaming eyes" at a wedding that is wedged somewhere between treatises in logic and lessons in Chinese.[30]) There may even be implications for how we conceive of ourselves: the fact that what ought (in virtue of its disvalue) to make a massive and traumatic impact on us leaves only a comparatively minor dent, while obviously advantageous, suggests that most of us lack the kind of emotional depth that accompanies deep insight into one's condition and which concomitantly enables deep suffering. We may thus begin to think of ourselves as less substantial, more superficial beings for our inability to hold on to our concern for great goods once we have lost them.

---

[27] See the interesting discussion by Shaun Nichols, "How Psychopaths Threaten Moral Rationalism," *The Monist*, LXXXV (2002): 285–304.

[28] There are many further issues here that I cannot pursue, for example, whether value judgments do not give rise to emotions rather than vice versa. My proposal should be viewed as a tentative attempt to identify the core issues, not to settle them. Some of these matters are explored at greater length in Michael Stocker's *Valuing Emotions* (New York: Cambridge, 1996), especially chapters 2–7, and Ronald de Sousa's *The Rationality of Emotion* (Cambridge: MIT, 1987).

[29] Proust, *In Search of Lost Time: Within a Budding Grove*, translated by C.K. Moncrieff et al. (New York: Modern Library), p. 340. Proust goes on rather implausibly to suggest that the reason we fear our prospective indifference is that this would indicate a "death of the self."

[30] W.V. Quine, *The Time of My Life* (Cambridge: MIT, 1985), p. 201.

Against all this, it might be said that our current pattern of resilience or at least something close to it is inescapable. "What on earth is the *alternative* to an emotional immune system that allows us to overcome trauma?" someone might demand. The point here is not (I hope) the trivial one that we might not be able to overcome entrenched features of human nature, but the deeper one that any imaginable alternative to resilience would make life unbearable. To modulate from the metaphor of immunity to clotting, the alternative simply sounds like a kind of emotional hemophilia. However, this objection involves a dialectical mistake. The goal here is not to advocate practical changes or even to show that resilience is bad *all things considered*. Perhaps the Adaptive Theory is true: the best (overall) response to grief is the one that promotes our interests the most, and perhaps our actual response is close to optimal in that respect. What I have been concerned to show, rather, is that there are *reasons* for regret, there are subtle and easily overlooked drawbacks to our partial immunity. Our embrace of the Adaptive Theory should be an uneasy one at best. That being the goal, there is no need to provide an alternative to resilience.

III

Not everyone yearns to be important to others or for supposedly profound insight into their life situation. Some readers may remain unmoved. For them I offer a different kind argument, designed to show that something like the concerns I have described *must* be valid. Consider a species of alien persons (or modified humans) who are like us except that members have *no* grief reactions at all to what would strike us as great tragedies. These are the *Super-resilient*. When their spouses drop dead in front of them, they shrug their shoulders and check what is on television. They remarry as soon as they are able to find another mate, often within weeks. They too deny not caring for their loved ones, in fact, investigation reveals that they are willing to walk through fire for their husbands and wives, and generally show tremendous concern before their loss. It is just that afterwards adaptive mechanisms operate so as immediately to extinguish any feelings of distress. The Super-resilient pity us for our insufficiently muted response to loss and wonder how we manage to endure through entire weeks of deep grief. Is there anything to be said against this extreme response to loss? Most people seem to think so. If that is right, this shows that the reasons for regretting resilience that I have been developing—or else others to the same effect—must at the very least get a grip on us, even if we insist that the Super-resilient merely embody a good thing taken too far. We cannot insist that there cannot be anything to regret about resilience if we find the Super-resilient troubling, though we might of course reject the particular reasons that were sketched above.

However, this argument may prompt a counterargument. The Super-resilient are at one extreme, but now consider the *Sub-resilient* at the other extreme. The Sub-resilient are like us except that they *never* cease caring as deeply for their spouses as at the moment of death, the loss of that relationship is as deeply felt at half a century as it is at half an hour. The bereaved Sub-resilient are consequently extremely unhappy people who feel they suffer from a kind of never-healing open wound. The undiminished depth of feeling for the deceased prevents them from remarrying. On the other hand, they have a profound sense of mutual importance since the loss of a spouse is so devastating. And they view us with a certain contempt, we seem to them incredibly shallow in our inability to realize the extent of our loss. The point now is not the recently discredited one that this alternative to resilience is unappealing—no one says otherwise—it is rather that the two extremes formed by the Sub- and Super-resilient bookend a continuum that includes a mean, and that we are at least roughly positioned as the proverbial healthy mean between the extremes. What is there to regret, what even counts against having a system of attitudes and emotional dispositions that avoids the pitfalls of either extreme and leaves us at the most livable point on the continuum, or at least reasonably close to it? Evolution may not be sensitive to factors like "mutual importance" or "insight" that were discussed earlier, but for whatever reason we nonetheless find ourselves endowed with a level of resilience that reflects these concerns to or close to the extent that we could endure before skirting the other extreme.

This seems to me the best response to the gloomy approach to love and death that I have been articulating. But alas, it too is unconvincing. Continua flanked by undesirable extremes are not all alike, sometimes what makes a center-value appealing is that it altogether avoids what bothers us about the extremes, but not always. Courage is the mean between cowardice and a kind of rash, wild disposition,[31] and objections to cowardice (or the other extreme) do not really apply to courage—it is not as if we reluctantly settle for courage while ruing our inability to access the benefits of cowardice. But take a more complex disposition not as easily assimilated to the Aristotelian model, say the beneficent treatment of people increasingly peripheral to us. On one extreme is the egalitarian disposition to treat total strangers and our children on an equal footing (either by elevating the former or demoting the latter or both). On the other side lies the tendency to narrow the scope of our beneficent concern to include only our spouse and children, say, and to treat even grandchildren, neighbors, and colleagues as we now do total strangers. Our actual disposition lies somewhere in between, though no doubt very far toward the latter end. Perhaps we are better off with our middle value than we would be under the egalitarian extreme, since we would then lack such goods as being special to our family members. But even if that is true all things considered, we still might think that there is something deeply troubling and unfortunate about our inability to recognize strangers as equally worthy objects of beneficence. We might even find the stance we are endowed with to be inconsistent with our considered moral beliefs.[32] And the same seems true of resilience. Quite apart from the question of whether evolution has instilled in us a level of resilience that is optimum relative to our various concerns, *any* deviation from the Sub-resilient raises difficult and painful questions. The kind of importance we have for one another, our irreplaceability or lack thereof, and our inability to appreciate fully our own condition as vulnerable victims of loss are issues that should concern us no matter how powerful the countervailing reasons are for accepting a mean (or more-nearly mean) position on the continuum. The values embraced by the Sub-resilient are much harder to dismiss than those associated with cowardice.

---

[31] For which we do not really have a name, as Aristotle points out (*Nicomachean Ethics* 1115b25).

We can now see why the sensationalist empirical findings we started out with are useful for motivating the discussion but ultimately inessential. Even if it turned out that we undergo intense distress for a few years rather than a few months, the question of whether we have reason to regret the attitudes toward a lost spouse that resilience instills would remain. (Though naturally as we begin to approach sub-resilience the issue begins to seem less pressing.) And, as I have argued, there are such reasons, though these reasons may be outweighed by other considerations, all things considered most of us might not wish to be significantly less resilient than we are.

---

[32] Of course, we might also take this stance as grounds for altering our considered moral beliefs, one man's *modus tollens* being another man's *modus ponens*.

For a philosopher to say any of this to the bereaved is a little like a spectator giving cheerful advice at a crucifixion But the marketplace of ideas is already full of such effronteries On the one hand, there is a widespread self-help approach to loss, which counsels therapeutic grief in order to begin what is called "the healing process,"[33] and on the other hand there is a technical literature, cited earlier, which emphasizes that many (though by no means all) victims of loss neither suffer much distress nor benefit in the long term from going through the supposed healing process But at this point it should be clear that there is something deeply wrong about both of these stances The first is invested in empirical assumptions that have simply proven false, and the second fails to register any of the profound complications we have seen resilience to involve Perhaps these complications I have reviewed should not be allowed to have the last word, given certain assumptions about the relative value of things like psychological well-being versus depth of insight into one's own condition, but that does not mean we should ignore them altogether

DAN MOLLER

---

[33] Though he would despise being associated with the self-helpers, C S Lewis's record of bereavement is perhaps the most impressive exemplar of their view, which emphasizes (a) initial anguish that (b) results in a certain catharsis or insight, followed (c) by a successful recovery and return to baseline that could not have been achieved without (a) and (b) See *A Grief Observed* (New York Harper Collins, 1961)

NEW BOOKS ANTHOLOGIES

PRIEST, GRAHAM, J C BEALL, and BRADLEY ARMOUR-GARB, eds *The Law of Non-Contradiction New Philosophical Essays* New York Oxford, 2004 xii, 443 p Cloth $85 00

PURI, BINDU, and HEIKO SIEVERS, eds *Reason Morality, and Beauty Essays on the Philosophy of Immanuel Kant* New York Oxford, 2007 xxi, 191 p Cloth $35 00

RADDEN, JENNIFER, ed *The Philosophy of Psychiatry A Companion* New York Oxford, 2004 xxi, 445 p Cloth $74 00

RAHE, PAUL A , ed *Machiavelli's Liberal Republican Legacy* New York Cambridge, 2006 lxii, 326 p Cloth $75 00

RAYO, AGUSTIN, and GABRIEL UZQUIANO, eds *Absolute Generality* New York Oxford, 2006 396 p Paper $45 00

REED, CHARLES, and DAVID RYALL, eds *The Price of Peace Just War in the Twenty-First Century* New York Cambridge, 2007 xvii, 340p Cloth $85 00, paper $29 99

REICHBERG, GREGORY M , HENRIK SYSE, and ENDE BEGBY, eds *The Ethics of War* Malden, MA Blackwell, 2006 xxii, 731 p Paper $54 95

REIDY, DAVID A , and MORTIMER N S SELLERS, eds *Universal Human Rights Moral Order in a Divided World.* Lanham, MD Rowman and Littlefield, 2005 viii, 238 p Cloth $75 00, paper $27 95

REIS, BURKHARD, ed *The Virtuous Life in Greek Ethics* New York Cambridge, 2006 x, 277 p Cloth $85 00

RIPSTEIN, ARTHUR, ed *Ronald Dworkin* New York Cambridge, 2007 x, 186 p Cloth $75 00, paper $24 99

ROBERTS, RODNEY C , ed *Injustice and Rectification* New York Peter Lang, 2005 191 p Cloth $54 95, paper $29 95

ROTHMAN, WILLIAM, ed *Cavell on Film.* Albany SUNY, 2005 xxvii, 399 p Cloth $89 50, paper $27 95

RUMBLE, W E , ed *British Philosophy of Law, 1832–1900, Editions 1–3* Thoemmes Continuum, 2005 720 p Cloth $450 00

RÜSEN, JÖRN, ed *Meaning and Representation in History* New York Berghahn, 2006 xiv, 269 p Cloth $80 00

RUTHERFORD, DONALD, ed *The Cambridge Companion to Early Modern Philosophy* New York Cambridge, 2006 xv, 421 p Cloth $75 00, paper $29 99

SALLES, RICARDO, ed *Metaphysics, Soul and Ethics in Ancient Thought Themes from the Work of Richard Sorabji.* New York Oxford, 2005 x, 592 p Cloth $115 00

SANTAS, GERASIMOS, ed *The Blackwell Guide to Plato's Republic* Malden, MA Blackwell, 2006 lx, 292 p Cloth $69 95, paper $29 95

SCARRE, CHRIS, and GEOFFREY SCARRE, eds *The Ethics of Archaeology Philosophical Perspectives on Archaeological Practice* New York Cambridge, 2006 xi, 318 p Cloth $80 00, paper $34 99

SCHAFER-LANDAU, RUSS, ed *Oxford Studies in Metaethics, Volume 1* New York Oxford, 2006 387 p Paper $35 00

SCHAFER-LANDAU, RUSS, and TERENCE CUNEO, eds *Foundations of Ethics An Anthology* Malden, MA Blackwell, 2007 xi, 504 p Paper $49 95

SCHNECK, STEPHEN F , ed *Letting Be Fred Dallmayr's Cosmopolitical Vision* Notre Dame University Press, 2006 x, 382 p Paper $35 00

SCHOTT, ROBIN MAY, ed *Feminist Philosophy and the Problem of Evil* Bloomington Indiana UP, 2007 325 p Paper $27 95

SCHRIFT, ALAN D , ed *Modernity and the Problem of Evil* Bloomington Indiana UP, 2005 200 p Cloth $50 00, paper $22 95

SEDLEY, DAVID, ed *Oxford Studies in Ancient Philosophy, Volume XXIX* New York Oxford, 2005 386 p Cloth $74 00, paper $49 95, *Oxford Studies in Ancient Philosophy, Volume XXX* New York Oxford, 2006 viii, 406 p Cloth $85 00, paper $45 00

SEESKIN, KENNETH, ed *The Cambridge Companion to Maimonides* New York Cambridge, 2006 xv, 406 p Cloth $70 00, paper $27 99

SELINGER, EVAN, and ROBERT P CREASE, eds *The Philosophy of Expertise* New York Columbia, 2006 vi, 421 p Cloth $49 50

SENNETT, JAMES F , and DOUGLAS GROOTHUIS, eds *In Defense of Natural Theology A Post-Humean Assessment* Downers Grove, IL InterVarsity Press, 2005 336 p Paper $26 00

SHANAHAN, TIMOTHY, ed *Philosophy 9/11 Thinking about the War on Terrorism* Chicago Open Court, 2005 xxi, 284 p Paper $29 95

SHAPIRO, STEWART, ed *The Oxford Handbook of Philosophy of Mathematics and Logic* New York Oxford, 2005 xv, 833 p Cloth $85 00

SHERWIN, SUSAN, and PETER K SCHOTCH, eds *Engaged Philosophy Essays in Honor of David Braybrooke* Toronto University Press, 2007 x, 425 p Cloth $65 00

SHOOK, JOHN R , and JOSEPH MARGOLIS, eds *A Companion to Pragmatism* Malden, MA Blackwell, 2006 xii, 431 p Cloth $124 95

SIMONS, MARGARET A , ed *The Philosophy of Simone de Beauvoir Critical Essays* Bloomington Indiana UP, 2006 xi, 316 p Cloth $65 00, paper $24 95

SINGER, PETER, ed *In Defense of Animals The Second Wave* Malden, MA Blackwell, 2006 x, 248 p Cloth $55 95, paper $23 95

SISTARE, CHRISTINE T , ed *Civility and Its Discontents Essays on Civil Virtue, Toleration, and Cultural Fragmentation* Lawrence Kansas UP, 2004 x, 310 p Cloth $39 95, paper $19 95

SMEYERS, PAUL, and MICHAEL A PETERS, eds *Postfoundationalist Themes in the Philosophy of Education Festschrift for James D Marshall* Malden, MA Blackwell, 2006 206 p Paper $34 95

SMITH, DAVID WOODRUFF, and AMIE L THOMASSON, eds *Phenomenology and Philosophy of Mind* New York Oxford, 2005 x, 322 p Cloth $95 00, paper $35 00

SMITH, JUSTIN E H , ed *The Problem of Animal Generation in Early Modern Philosophy* New York Cambridge, 2006 xiii, 456 p Cloth $85 00

SOBER, ELLIOTT, ed *Conceptual Issues in Evolutionary Biology, Third Edition* Cambridge MIT, 2007 xxviii, 612 p Paper $42 00

SORELL, TOM, and G A J ROGERS, eds *Analytic Philosophy and History of Philosophy* New York, Oxford 2005 229 p Cloth $55 00

SOUTH, JAMES B , and JACOB M HELD, eds *James Bond and Philosophy Questions Are Forever* Chicago Open Court, 2006 xv, 244 p Cloth $17 95

STAINTON, ROBERT J , ed *Contemporary Debates in Cognitive Science* Malden, MA Blackwell, 2006 xiv, 339 p Cloth $84 95, paper $34 95

STERN, DAVID G , and BÉLA SZABADOS, eds *Wittgenstein Reads Weininger* New York Cambridge, 2004 vii, 197 p Cloth $70 00, paper $24 99

STERNBERG, ROBERT J , and JENNIFER JORDAN, eds *A Handbook of Wisdom Psychological Perspectives* New York Cambridge, 2005 xvii, 387 p Cloth $75 00, paper $34 99

STRANGE, STEVEN K , and JACK ZUPKO, eds *Stoicism Traditions and Transformations* New York Cambridge, 2004 xi, 295 p Cloth $70 00

SURBER, JERE O'NEIL, ed *Hegel and Language* Albany SUNY, 2006 vii, 259 p Cloth $75 00

SZABADOS, BÉLA, and ELDON SOIFER, eds *Hypocrisy Ethical Investigations* Orchard Park, NY Broadview, 2004 352 p Paper $19 95

TALIAFERRO, CHARLES, and ALISON J TEPLY, eds *Cambridge Platonist Spirituality* New York Paulist, 2004 xii, 233 p Paper $34 95

TAYLOR, JAMES STACEY, ed *Personal Autonomy New Essays on Personal Autonomy and Its Role in Contemporary Moral Philosophy* New York Cambridge, 2005 ix, 350 p Cloth $70 00

THOMASSEN, LASSE, ed *The Derrida-Habermas Reader* Chicago University Press, 2006 viii, 321 p Cloth $73 00, paper $29 00

THOMSON, JUDITH, and ALEX BYRNE, eds *Content and Modality Themes from the Philosophy of Robert Stalnaker* New York Oxford, 2006 viii, 304 p Cloth $74 00

TIMMONS, MARK, ed *Disputed Moral Issues A Reader* New York Oxford, 2007 xii, 564 p Paper $74 00

TRAIGER, SAUL, ed *The Blackwell Guide to Hume's Treatise* Malden, MA Blackwell, 2006 xi, 307 p Cloth $74 95, paper $29 95

TREMAIN, SHELLEY, ed *Foucault and the Government of Disability* Ann Arbor Michigan UP, 2005 xvii, 340 p Cloth $60 00, paper $27 95

TRUESWELL, JOHN C , and MICHAEL K TANNENHAUS, eds *Approaches to Studying World-Situated Language Use* Cambridge MIT, 2005 xxii, 379 p Paper $35 00

TURNER, MARK, ed *The Artful Mind Cognitive Science and the Riddle of Human Creativity* New York Oxford, 2006 xvi, 314 p Cloth $35 00

URBINATI, NADIA, and ALEX ZAKARAS, eds *J S Mill's Political Thought A Bicentennial Reassessment* New York Cambridge, 2007 viii, 392 p Cloth $85 00, paper 27 99

VALLS, ANDREW, ed *Race and Racism in Modern Philosophy* Ithaca, NY Cornell, 2005 293 p Cloth $45 00, paper $21 95

VAN DEN BRINK, BERT, and DAVID OWEN, eds *Recognition and Power Axel Honneth and the Tradition of Critical Social Theory* xiv, 399 p Cloth $85 00

VANHOOZER, KEVIN J , JAMES K A SMITH, and BRUCE ELLIS BENSON, eds *Hermeneutics at the Crossroads* Bloomington Indiana UP, 2006 xviii, 246 p Cloth $65 00, paper $24 95

VAN INWAGEN, PETER, ed *Christian Faith and the Problem of Evil* Grand Rapids, MI Eerdmans, 2004 xiv, 316 p Paper $35 00

VAN INWAGEN, PETER, and DEAN ZIMMERMAN, eds *Persons Human and Desire* New York Oxford, 2007 vii, 375 p Paper $35 00

WAINWRIGHT, WILLIAM J , ed *The Oxford Handbook of Philosophy of Religion* New York Oxford, 2005 x, 550 p Cloth $74,00

WALKER, REBECCA L , and PHILIP J IVANHOE, eds *Working Virtue Virtue Ethics and Contemporary Moral Problems* New York Oxford, 2007 ix, 319 Cloth $85 00

WALLACE, R JAY, et al , eds *Reason and Value Themes from the Moral Philosophy of Joseph Raz* New York Oxford, 2004 xi, 429 p Cloth $74 00

WARTENBERG, THOMAS E , and ANGELA CURRAN, eds *The Philosophy of Film Introductory Text and Readings* Malden, MA Blackwell, 2005 xii, 308 p Paper $34 95

WASSERMAN, DAVID, JEROME BICKENBACH, and ROBERT WACHBROIT, eds *Quality of Life and Human Difference* New York Cambridge, 2005 xi, 273 p Cloth $70 00, paper $24 99

WELCH, CHERYL B , ed *The Cambridge Companion to Tocqueville* New York Cambridge, 2006 xxvii, 428 p Cloth $80 00, paper $29 99

WELCHMAN, JENNIFER, ed *The Practice of Virtue Classic qnd Contemporary Readings in Virtue Ethics* Indianapolis Hackett, 2006 xxv, 224 p Cloth $37 95, paper $12 95

WEST, HENRY R , ed *The Blackwell Guide to Mill's Utilitarianism* Malden, MA Blackwell, 2006 ix, 275 p Cloth $74 95, paper $29 95

WILLIAMS, MELISSA, and STEPHEN MACEDO, eds *Political Exclusion and Domination* New York University Press, 2005 xiv, 368 p Cloth $55 00

WILLIAMS, MEREDITH, ed *Wittgenstein's Philosophical Investigations* Lanham, MD Rowman and Littlefield, 2007 ix, 238 p Cloth $75 00, paper $29 95

WINKLER, KENNETH P , ed *The Cambridge Companion to Berkeley* New York Cambridge, 2005 xiv, 454 p Cloth $75 00, paper $29 99

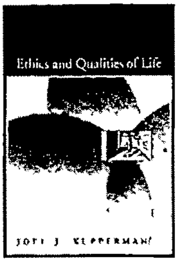
WOLF, MICHAEL P , and MARK NORRIS LANCE, ed *The Self-Correcting Enterprise Essays on Wilfrid Sellars* New York Rodopi, 2006 274 p Cloth $81 00

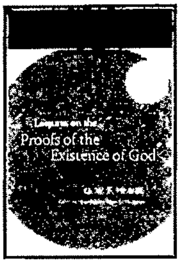
WRATHALL, MARK A , ed *U2 and Philosophy How to Decipher an Atomic Bomb* Chicago Open Court, 2006 xvi, 228 p Paper $17 95

WRIGHT, ERIK OLIN, ed *Redesigning Distribution Basic Income and Stakeholder Grants as Cornerstones for an Egalitarian Capitalism.* New York Verso, 2006 xii, 228 p Cloth $65 00, paper $25 00

ZIMMERMAN, DEAN W , ed *Oxford Studies in Metaphysics, Volume 2* New York Oxford, 2006 x, 400 p Cloth $99 00, paper $35 00

New from OXFORD





Now in paperback!

Deeper than Reason
Emotion and its Role in Literature, Music, and Art
JENEFER ROBINSON

"Robinson is the only philosopher of music known to me who seems to have understood that musical experience involves the activation of the motor system and to have grasped this fact's fundamental importance *Deeper than Reason* is impressive Indeed, it is the best philosophical monograph I know on emotion and the arts Among its many achievements is one most rare it manages to be both accessible and profound "
—*Modern Philology*

2007 • 515 pp • paper • $35 00

Hegel
Lectures on the Proofs of the Existence of God
Edited and translated by
PETER C HODGSON

Peter C Hodgson provides a new translation of Hegel's 1829 lectures on the proofs of the existence of God, based on the definitive German edition Coming late in his career, these lectures give us the great philosopher's final and most seasoned thinking on a topic of obvious significance to him, that of the reality status of God and ways of knowing God

(Hegel Lectures)
2007 • 224 pp • $90 00

Ethics and Qualities of Life
JOEL J KUPPERMAN

Ethics and Qualities of Life looks at what enters into ethical judgment and choice Interpretation of a case and of what the options are is always a factor, as is a sense of the possible values at stake

2007 • 206 pp • $45 00

The Possibility of Knowledge
QUASSIM CASSAM

How is knowledge of the external world possible? How is knowledge of other minds possible? How is a priori knowledge possible? These are all examples of how-possible questions in epistemology In this highly original book, Quassim Cassam explains how such questions arise and how they should be answered

2007 • 256 pp • $49 95

Ordinary Objects
AMIE L THOMASSON

Ordinary Objects shows how to develop a common-sense ontology and defend it against a variety of recent eliminativist arguments It argues that the apparently diverse eliminativist arguments rest on a few shared assumptions, and that questioning these assumptions gives us reason to reevaluate the proper methods and limits of metaphysics

2007 • 256 pp • $55 00

OXFORD
UNIVERSITY PRESS

Prices are subject to change and apply only in the US To order, please call 1-800-451-7556 In Canada, call 1-800-387-8020 Visit our website at www oup com/us

New from OXFORD

David Hume
A Treatise of Human Nature

Edited by DAVID FATE NORTON
and MARY J NORTON

David and Mary Norton present the definitive scholarly edition of Hume's *Treatise*, one of the greatest philosophical works ever written

(The Clarendon Edition of the Works of David Hume)

Volume 1 Texts 2007 464 pp $99 00

Volume 2 Editorial Material 2007 624 pp $135 00

2 Volume Set 2007 1090 pp $199 00

Reason of State, Propaganda and the Thirty Years' War
An Unknown Translation by Thomas Hobbes

NOEL MALCOLM

Acclaimed writer and historian Noel Malcolm presents his sensational discovery of a new work by Thomas Hobbes (1588-1679) a propaganda pamphlet on behalf of the Habsburg side in the Thirty Years' War, translated by Hobbes from a Latin original

2007 240 pp 2 b/w photographs $65 00

Truth and Ontology
TRENTON MERRICKS

A central question in philosophy is whether and how truth depends on the world Trenton Merricks looks at how answers to this question bear on a variety of philosophical debates

2007 206 pp $49 95

Simple Sentences
JENNIFER M SAUL

The phenomenon of substitution failure is a longstanding focus of discussion for philosophers of language Jennifer Saul presents the first full-length treatment of this puzzling feature of language, and explores its implications for other theories of reference and names, and for the methodology of semantics

2007 176 pp $45 00

Content, Cognition, and Communication
Philosophical Papers II

NATHAN SALMON

Nathan Salmon presents a selection of nineteen of his essays from the early 1980s to 2006, on a set of closely connected topics central to analytic philosophy

2007 420 pp cloth $120 00 paper $45 00

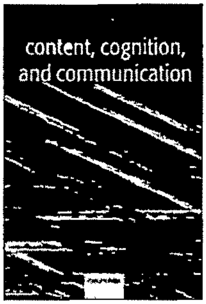
The Oxford Handbook of Ethical Theory

Edited by DAVID COPP

The Oxford Handbook of Ethical Theory is a major new reference work in ethical theory consisting of commissioned essays by leading moral philosophers.

(Oxford Handbooks in Philosophy)

2007 680 pp. cloth $74.00 paper $39.95



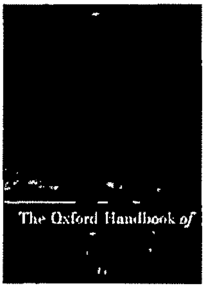
The Oxford Handbook of Philosophy of Mathematics and Logic

Edited by STEWART SHAPIRO

"*The Oxford Handbook of the Philosophy of Mathematics and Logic* is a very accessible, wide ranging work that serves not only to indicate the 'state of the art' in the given area, but, remarkably, also serves as a very fine introduction to the field. I recommend it highly, both to workers in the given field and, equally, to the 'general philosopher,' regardless of one's main area."
—Notre Dame Philosophical Reviews

(Oxford Handbooks in Philosophy)

2007 832 pp., 9 line illus.
cloth $90.00 paper $39.95



The Oxford Handbook of Philosophy of Religion

Edited by WILLIAM WAINWRIGHT

"An excellent reference source that makes a significant contribution to the discipline. Written by major scholars in the field, the chapters include not only descriptive information but also critical and interpretive insights. Highly recommended."—CHOICE

(Oxford Handbooks in Philosophy)

2007 528 pp. cloth $74.00 paper $29.95

Oxford Studies in Metaethics

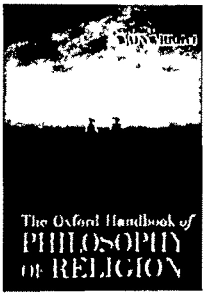
Volume II

RUSS SHAFER-LANDAU

Oxford Studies in Metaethics is the only periodical publication devoted exclusively to original philosophical work on the foundations of ethics. Its broad purview includes work being done at the intersections of ethical theory with metaphysics, epistemology, philosophy of language, and philosophy of mind.

(Oxford Studies in Metaethics)

2007 280 pp. cloth $99.00 paper $34.00



OXFORD
UNIVERSITY PRESS

To order, call 1-800-451-7556. In Canada, 1-800-387-8020.
Visit our web site at www.oup.com/us

CAMBRIDGE

Outstanding Titles in Philosophy

Cambridge Introductions to Philosophy

An Introduction to Political Philosophy
Colin Bird
$80.00 Hb 978-0-521-83625-8
$24.99 Pb 978-0-521-54482-5 322 pp

An Introduction to the Philosophy of Language
Michael Morris
$75.00 Hb 978-0-521-84215-0
$29.99 Pb 978-0-521-60311-9 336 pp

An Introduction to the Theory of Knowledge
Noah Lemos
$85.00 Hb 978-0-521-84213-6
$24.99 Pb 978-0-521-60309-6 242 pp

An Introduction to Godel's Theorems
Peter Smith
$85.00 Hb 978-0-521-85784-0
$29.99 Pb 978-0-521-67453-9 376 pp

Prices subject to change

www.cambridge.org/us

CAMBRIDGE
UNIVERSITY PRESS

CAMBRIDGE

Outstanding Titles in Philosophy

Cambridge Texts in the History of Philosophy

Hume *An Enquiry concerning Human Understanding*
And Other Writings
Edited by Stephen Buckle
$50 00 Hb 978 0 521 84340 9
$22 99 Pb 978 0 521 60403 1
278 pp

Hume *Dialogues Concerning Natural Religion*
And Other Writings
Edited by Dorothy Coleman
$55 00 Hb 978 0 521 84260 0
$19 99 Pb 978 0 521 60359 1
310 pp

Neurophilosophy at Work
Paul Churchland
$75 00 Hb 978 0 521 86472 5
$24 99 Pb 978 0 521 69200 7
262 pp

Ideas in Context

Locke, Language and Early-Modern Philosophy
Hannah Dawson
$90 00 Hb 978 0 521 85271 5
378 pp

Roman Monarchy and the Renaissance Prince
Peter Stacey
$90 00 Hb 978 0-521 86989 8
360 pp

Explaining Social Behavior
More Nuts and Bolts for the Social Sciences
Jon Elster
$85 00 Hb 978 0 521-77179 5
$27 99 Pb 978 0 521 77744 5
496 pp

Heidegger's *Being and Time*
An Introduction
Paul Gorner
Cambridge Introductions to Key Philosophical Texts
$80 00 Hb 978-0 521 83322-6
$23 99 Pb 978 0 521 54072 8
204 pp

Rethinking Informed Consent in Bioethics
Neil C Manson and Onora O'Neill
$85 00 Hb 978 0-521-87458 8
$32 99 Pb 978 0 521-69747-7
226 pp

The Cambridge Companion to Locke's *Essay Concerning Human Understanding*
Edited by Lex Newman
Cambridge Companions to Philosophy
$80 00 Hb 978 0 521 83433 9
$29 99 Pb 978 0 521 54225 8
502 pp

The Price of Peace
Just War in the Twenty-First Century
Edited by Charles Reed and David Ryall
Foreword by Richard Dannatt
$85 00 Hb 978 0-521 86051 2
29 99 Pb 978 0 521 67785 1
358 pp

Ronald Dworkin
Arthur Ripstein
Contemporary Philosophy in Focus
$75 00 Hb 978 0 521 66289 5
$24 99 Pb 978 0 521 66412 7
192 pp

Group Rationality in Scientific Research
Husain Sarkar
$80 00 Hb 978 0 521 87113 6
300 pp

Democracy and Legal Change
Melissa Schwartzberg
Cambridge Studies in the Theory of Democracy
$80 00 Hb 978 0 521-86652 1
240 pp

Remastering Morals with Aristotle and Confucius
May Sim
$85 00 Hb 978 0 521 87093-1
138 pp

Now in Paperback!

Ayn Rand's Normative Ethics
The Virtuous Egoist
Tara Smith
$25 99 Pb 978 0 521-70546 2
328 pp

Social Philosophy after Adorno
Lambert Zuidervaart
$80 00 Hb 978 0 521 87027 6
$25 99 Pb 978 0-521 69038 6
232 pp

Recognition and Power
Axel Honneth and the Tradition of Critical Social Theory
Bert van den Brink and David Owen
$85 00 Hb 978 0 521 86445 9
414 pp

Prices subject to change

www cambridge org/us

CAMBRIDGE
UNIVERSITY PRESS

New books from The MIT Press

Language, Consciousness, Culture

Essays on Mental Structure

Ray Jackendoff

An integrative approach to human cognition that encompasses the domains of language, consciousness, action, social cognition, and theory of mind that will foster cross-disciplinary conversation among linguists, philosophers, psycholinguists, neuroscientists, cognitive anthropologists, and evolutionary psychologists

Jean Nicod Lecture series • A Bradford Book
710 pp , 21 illus $36 cloth

Feeling Pain and Being in Pain

Second Edition

Nikola Grahek

foreword by Daniel Dennett

"Nikola Grahek had both the conceptual acuity of a well-trained philosopher and an intimate, broad-based knowledge of the empirical science relating to his chosen topic, pain Putting the two together, he came up with some startling and attention-shifting proposals about how to make sense of pain, revisionary proposals that should clarify the thinking of scientist and layman alike "
— from the foreword by Daniel Dennett

A Bradford Book • 168 pp $30 cloth

Causation and Explanation

edited by Joseph Keim Campbell, Michael O'Rourke, and Harry S Silverstein

Leading scholars discuss the development and application of theories of causation and explanation, offering a state-of-the-art view of current work on these two topics

Topics in Contemporary Philosophy, volume IV
480 pp , 66 illus $36 paper

Conditionals

Nicholas Rescher

A unified treatment of conditionals based on epistemological principles rather than the semantical principles in vogue over recent decades

A Bradford Book • 272 pp , 15 illus $36 cloth

The Letters of George Santayana

Book Seven, 1941-1947

edited and with an introduction by William G Holzberger

The seventh and penultimate book of the letters of American philosopher George Santayana, covering the years 1941 to 1947 and including letters to such correspondents as Daniel Cory, John Hall Wheelock, Robert Lowell, and others

648 pp $70 cloth

Zone Books

Nongovernmental Politics

edited by Donald Feher

The past, present, and future prospects of nongovernmental politics—political activism that eschews the question of *who* should govern to better focus on that of *how* government is exercised

Distributed for Zone Books • 672 pp , 100 color illus , 150 black & white illus $39 95 paper

Who Are You?

Identification, Deception, and Surveillance in Early Modern Europe

Valentin Groebner

translated by Mark Kyburz and John Peck

The prehistory of modern passport and identification technologies the documents, seals, and stamps, that could document and transform their owner's identity

Distributed for Zone Books
340 pp , 22 illus $35 cloth

The Inner Touch

Archaeology of a Sensation

Daniel Heller-Roazen

An original, elegant, and far-reaching philosophical inquiry into the sense of being sentient—what it means to feel that one is alive—that draws on philosophical, literary, psychological, and medical accounts from ancient, medieval, and modern cultures life

Distributed for Zone Books • 300 pp $33 cloth

To order call **800-405-1619**
http //mitpress mit edu

Without Justification

Jonathan Sutton

An argument that takes issue with the contemporary epistemological consensus that justification is distinct from knowledge, proposing instead that justified belief simply is knowledge, and arguing in detail that a belief is justified when and only when it amounts to knowledge

A Bradford Book • 208 pp $30 paper

Wittgenstein and the Moral Life

Essays in Honor of Cora Diamond

edited by Alice Crary

"The distinguished philosophers Alice Crary has assembled in this fine book pay fitting tribute to one of the world's most important living philosophers "
— Raimond Gaita, King's College London

Representations and the Mind series
432 pp $36 paper

Reliable Reasoning

Induction and Statistical Learning Theory

Gilbert Harman and Sanjeev Kulkarni

The implications for philosophy and cognitive science of developments in statistical learning theory

Jean Nicod Lecture series • A Bradford Book
114 pp , 19 illus $30 cloth

now in paperback

Panpsychism in the West

David Skrbina

"Provides a long overdue and much needed reexamination of this age-old doctrine, one which still retains some fascination for modern philosophy of mind "
— William Seager, University of Toronto at Scarborough

336 pp $21 paper

now in paperback

Social Empiricism

Miriam Solomon

"Not only important but short, clear, lively, and well-supported by case studies will serve well both philosophers and their students " — Catherine Hundleby, *Philosophy in Review*

A Bradford Book • 196 pp $18 paper

Democracy across Borders

From Dêmos to Dêmoi

James Bohman

"This book will become a standard point of reference for debates on key issues in the theory and practice of international democratic politics " — John S Dryzek, Australian National University

Studies in Contemporary German Social Thought series • 232 pp $35 cloth

now in paperback

Neither Brain nor Ghost

A Nondualist Alternative to the Mind-Brain Identity Theory

W Teed Rockwell

"Using a powerful combination of Dewey's pragmatism and dynamical systems theory, [Rockwell] proposes a bold alternative to Cartesian materialism that deserves careful scrutiny "
— J A Scott Kelso, Florida Atlantic University

A Bradford Book • 256 pp , 6 illus $18 paper

now in paperback

Reconstructing the Cognitive World

The Next Step

Michael Wheeler

"This book has the potential to radically reshape how cognitive science and analytical philosophy view the mind "
— Mark A Bedau, Reed College, Editor-in-Chief of *Artificial Life*

A Bradford Book • 432 pp $20 paper

now in paperback

Molecular Models of Life

Philosophical Papers on Molecular Biology

Sahotra Sarkar

"Sarkar stands out among philosophers of science for his ability to combine conceptual, historical, and technical considerations in compelling perspectives on the philosophical questions raised by contemporary biology " — Paul E Griffiths, University of Pittsburgh

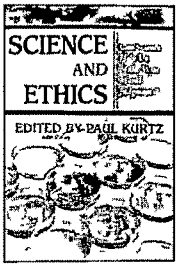
352 pp $25 paper

SCIENCE AND ETHICS

Can Science Help Us Make Wise Moral Judgments?

Edited by Paul Kurtz, *with the assistance of David Koepsell*

350 pp • ISBN 978-1-59102-537-5 • PB • $21



A unique collection of authors who maintain that science can help us make wise choices and that an increase in scientific knowledge can help modify our ethical values and bring new ethical principles into social awareness.

PUTTING PHILOSOPHY TO WORK

Inquiry and Its Place in Culture—*Essays on Science, Religion, Law, Literature, and Life*

Susan Haack

336 pp • ISBN 978-1-59102-528-3 • HC • $32

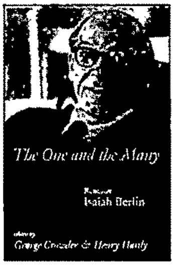
This wide-ranging collection of Haack's new and recent essays offers applied philosophy in a new key and a fresh voice. Whether she is telling the "sad, sleazy saga" of a drug company's efforts to suppress the unfavorable results of a clinical trial it had funded, disentangling the many senses of "coherent," or reflecting on the growth of intellectual character, she writes with the clarity, verve, and wit her readers have come to expect.

THE ONE AND THE MANY

Reading Isaiah Berlin

Edited by George Crowder and Henry Hardy

300 pp • ISBN 978-1-59102-448-4 • HC • $28



"[A] fitting tribute to Berlin. It is a multidimensional exploration of the man and the ideas: his Judaism, his moderate nationalism, his ambivalence about the Enlightenment, his reaction to Stalinism, and a great deal more. It is a rich, original, and nourishing volume."

—Alan Ryan, warden of New College, Oxford

SKEPTICISM

An Anthology

Edited by Richard H. Popkin and José R. Maia Neto

350 pp • ISBN 978-1-59102-474-3 • PB • $29

This anthology contains the principal texts of the Western skeptical tradition from its origins in antiquity to contemporary philosophy. An introduction on the origins, kinds, and significance of philosophical skepticism puts the various readings into historical context.

DENNETT AND RICOEUR ON THE NARRATIVE SELF

Joan McCarthy

220 pp • ISBN 978-1-59102-548-1 • PB • $39.95

Is the self some kind of real entity or a very convincing illusion? Drawing on the work of two leading philosophers, Daniel Dennett and Paul Ricoeur, McCarthy examines how each of these thinkers casts the self in narrative terms.

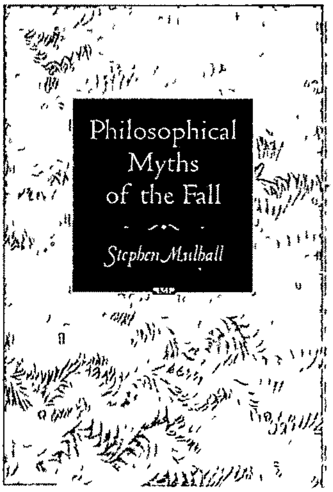
Prometheus Books
Toll Free: 800-421-0351 / Fax: 716-691-0137
www.prometheusbooks.com

Nietzsche's Political Skepticism

Tamsin Shaw

"A very fine piece of research and argument. This book reads Nietzsche against the grain, since it challenges both a fair bit of what Nietzsche explicitly says about his own beliefs as well as the interpretations that these pronouncements have inspired. But it does so successfully, for Shaw questions those pronouncements in the name of what is probably Nietzsche's strongest drive—his relentless truth-seeking."
—Bernard Yack,
Brandeis University

Cloth $24.95 978-0-691-13322-5 Due October



New in paperback

Philosophical Myths of the Fall

Stephen Mulhall

"Mulhall's re-opening of issues of fall and redemption is not so much a re-construction of a specific answer as it is a re-articulation of the germane questions and a re-thinking of possible responses. Mulhall simultaneously puts into question the adequacy of the philosophical myths and raises the prospect of reconsidering the Christian account."
—Thomas Hibbs, *First Things*

Princeton Monographs in Philosophy
Harry G. Frankfurt, Series Editor
Paper $19.95 978-0-691-13392-8

PRINCETON UNIVERSITY PRESS
800.777.4726
press.princeton.edu

ANNOUNCING:

THE JOURNAL OF PHILOSOPHY

ONLINE

The Journal of Philosophy has added
a print and online subscription option for institutions

SUBSCRIPTION OPTIONS

Libraries and Institutions (print and online)	$180 00
Libraries and Institutions (print only)	$100 00
Individuals	$45 00
Students, retired/unemployed philosophers	$20 00
Postage outside the U S	$15 00

Please consult our website at www journalofphilosophy org
for more details